

Negotiation by stops and starts

The proposed bilateral negotiations on space weapons, planned for September in Vienna, are now in doubt. But they should best be postponed until after the US election.

THE United States and the Soviet Union seem bent on setting new standards of folly in their bilateral negotiations on arms control. By now the world is used to negotiations that are abandoned after months of patient effort. But at the weekend, a novel method of mutual recrimination seemed on the cards. It is now almost certain that the negotiations on space weapons arranged for 18 September at Vienna will be abandoned before they have begun. Is this a record?

The omens for Vienna have never been encouraging. The Soviet invitation to the talks at the end of June was intended as a challenge. To everybody's surprise, it was promptly accepted by the United States, which then angered the Soviet Union by saying that it would also raise at Vienna the questions left hanging since the abandonment of the Geneva negotiations on ballistic missiles. That, the Soviet Union said, was a precondition, and unacceptable, whereupon the argument between the two governments shifted to the form of words in which a preliminary statement about the objectives of the Vienna talks should be made public. By asking that there should be a moratorium on the testing of space weapons while the talks were in progress, the Soviet Union appeared to be imposing preconditions of a kind that it had earlier declared unacceptable. Precisely why the draft statement prepared by the United States gave such offence in Moscow is not yet clear. What does stand out, however, is that neither side is adequately prepared for negotiations. By seeking credit for their willingness to be seen talking, without knowing what they would talk about, they have jeopardized the chance that serious negotiations on a serious agenda will take place. For that, the superpowers are equally to blame.

The need now is for composure. With all the disappointment there will be, in Europe for example, that the Vienna negotiations may not begin, it may be forgotten that the Soviet invitation is not some daring innovation but, rather, a proposal that the two superpowers should resume the series of exploratory talks on space weapons begun at Vienna in 1978 but left on ice the following year, after Western discontent with events in Poland and Afghanistan. Since then, not much has happened. The Soviet Union has published (in 1981 and 1983), draft treaties to prohibit the testing and deployment of anti-satellite weapons (ASATs), and the United States has explained that the texts are unacceptable because their proposed constraints cannot be verified. In the past year, the role of space weapons in strategic conflict has been further dramatized by President Ronald Reagan's star-wars speech, on 23 March 1983. It is almost certainly true that the goal of a nearly perfect defence against hostile ballistic missiles will never be attainable (see p.353), but the US Administration is also on safe ground in saying that the Strategic Defense Initiative (SDI) will entail research and development certain to yield novel weapons systems of great military significance.

Objectives

What, in these circumstances, should be the objectives of serious negotiations? The superpowers are already bound by two important restraints — the treaty on the Peaceful Uses of Outer Space (of which the Peoples' Republic of China is now, fortunately, a signatory) and the Anti-Ballistic Missile (ABM) treaty, both of them signed in 1972. The first prohibits the deployment of "weapons of mass destruction" (understood as

nuclear bombs) in space, and requires governments to follow the spirit of international law in their activities beyond the atmosphere (which is presumably, among other things, an interdiction of piracy). The second (as amended in 1974) limits the deployment of ABM systems to one on each side, and prohibits not merely the deployment and testing but even the development of anti-ballistic missile systems that are not based on land.

So far, the two treaties appear to have been successful. Neither side interfered with an Earth satellite belonging to the other. The United States complains that the huge radar system being built at Krasnayarsk in Central Siberia represents a Soviet violation of the ABM treaty, but the circumstances are obscure. What is, however, clear is that it cannot be very long before the research and development programme being mounted in the United States under the banner of SDI bumps head-on into the provisions of the ABM treaty. A ban on ASAT weapons, the most obvious goal of the Vienna talks, would bring about that conflict sooner. It follows that the Soviet Union cannot reasonably ask the United States to sign on some simple dotted line without conceding that broader issues must also be considered.

The argument goes like this. ASATs represent a danger for both sides. Because they can in principle destroy military reconnaissance satellites, they raise the awesome prospect that a pre-emptive attack might be launched by one side on the other without warning, making retaliation less certain. So as things are, both sides would sleep more easily at night if ASATs were banned. One practical difficulty is that the airborne system demonstrated earlier this year by the United States, based on the F-15 aircraft, is more flexible than the Soviet rocket-based system, and more easily concealed. There is also the problem that early warning satellites may soon be most easily rendered useless electronically, without being physically destroyed. So while the present time may be the last opportunity for nipping these emerging technologies in the bud, a secure treaty will require a fuller exchange of information about the purposes of military experiments beyond the atmosphere than would have been necessary a few years ago. Even a simple ASAT treaty would require a degree of disclosure unprecedented in bilateral treaties.

Disaster

There is a further complication. Mistaken though the star wars speech may have been, the programme of research and development on which the United States Administration is embarking should at some stage yield a capacity to destroy at least some proportion of a hostile missile force. In the absence of any further agreement on the military uses of space, the United States would no doubt be compelled, a few years from now, to exercise its right to withdraw from the ABM treaty on the grounds that its "supreme interests" are jeopardized by its inability to defend itself against hostile missiles. That turn of events, which would undermine confidence in all existing agreements on arms control, would be a disaster for everybody. One result would no doubt be a further deployment of nuclear missiles to counteract the knowledge that some of them might be destroyed in flight. But from that it follows that, if the development of more effective ABM systems is to be prevented, there should be complementary agreements on the strength of each side's strategic missile force, preferably much less than that allowed by the still-unratified



SALT II treaty. To that extent, the United States is right in asking that missiles should be considered alongside ASATs.

Whether such a broad agenda can ever be agreed is anybody's guess. A quick start on negotiation, which would be of symbolic value, is less important than the need that the ground should be prepared thoroughly in advance. And there is much to be said for postponing negotiations until after the US elections in November. President Reagan makes no secret of his eagerness to be seen to be talking to the Soviet Union during the pre-election period, but no doubt has also calculated that it would be an electoral handicap if talks began and then broke down in a flurry of recrimination.

Either way, it is unthinkable that negotiations conducted in parallel with an election campaign could yield the best result. But it should also serve the interests of the Soviet Union to wait until after November, if only in case there is a change of administration in the United States. Then, at least the negotiators should know with and for whom they are negotiating. □

Secret science

Governmental efforts to impose spurious secrecy on the scientific enterprise can be dangerous.

It is a safe bet that many official secrets are neither official nor secret. Government secrecy, however well justified when first imposed, has a way of becoming an end in itself. At no time has this been more apparent than during the Reagan Administration, ironically elected on the pledge to curtail government usurpation of individual freedom. Thus the Reagan Administration has imposed unprecedented restrictions on contacts with the press by government officials, has sought to impose lifelong censorship on government employees with access to classified materials and has even tried to vitiate the Freedom of Information Act.

Tales of high government silliness about secrecy are, of course, nothing new. Generals' laundry lists have become top secret. But it is more than just silliness when secrecy becomes a veil behind which travesties of public policy are committed. One such is the way in which US intelligence officials have arrogated to themselves responsibilities they could never claim in the open by seeking to bridge a supposed strategic gap between the United States and the Soviet Union — in parapsychology. Last February, presidential science adviser George Keyworth all but acknowledged the existence of a parapsychology development programme. In response to a question on the subject, he paused and then said, "Let me say this only, and that is that in our pursuit of research in the really critical areas of military technology, we do our very best to let our imaginations and our creativity be as effective as possible. We pay attention to Soviet programs, and no areas are going to be clearly rejected on parochial grounds."

Nobody knows exactly what the Central Intelligence Agency and the Department of Defense may be up to, which is why attention has been concentrated on the supposedly investigative reporting of one Don McRae, who has just published an account of the US Government's interest in parapsychology (*Mind Wars: The true story of secret government research into the military potential of psychic weapons*, St Martin's Press, New York, 1984). McRae's story in any case makes entertaining reading. We learn, for example, that the US Navy hired 34 psychics to keep track of Soviet submarines, that the National Security Agency is applying clairvoyance to code-breaking and that worries over Soviet psychics peering into concrete silos led to the abandonment of the "race-track" basing mode for the MX missile. Only the untested arrogance that secrecy affords could permit such thinking to go on inside the government. Even a little dose of peer review (or scientific expertise) would do wonders to cure it.

The same untested pseudo-science seems to be behind the administration's plans to expand the use of the polygraph or "lie detector", which seems to flourish in the climate of national security and darkness of secrecy.

The one Department of Defense official who tried to shed the light of science on the plans is said to have been given a dressing-down from General Richard Stilwell, Deputy Under-Secretary of

Defense and a retired four-star general who is not used to being told he is wrong. The official finally gave up, left the department and now says he should have known better than to try to talk science to people who wear copper bracelets to cure their arthritis. It will be interesting to see whether those hoping to keep polygraphs out of the British Government's Communications Headquarters in Cheltenham will be any more successful. □

Consistency in France

The new French Government may be even more zealous in the application of science.

NOT all of France would agree with the opinion that President François Mitterrand is the epitome of consistency, but there is a good deal of support for that proposition. Outwardly, the evidence is stacked the other way. Since his election in 1981, the President has blown hot and cold on the principal issues raised during his election. To begin with, he was faithful to the promise that France would be run on socialist lines, and embarked on a programme of nationalization as remarkable for its diversity as for its scale. At the outset, the government of France clearly believed that it would be possible to banish many social problems, unemployment for example, by deficit financing contained within the bounds of plausibility only by draconian taxes. By the spring of 1983, however, it was clear that the recipe would lead to economic disaster, so the President and the government changed economic course, settling instead for the period of retrenchment that has produced the past year's economic privations in Lorraine and other regions of France best-known for their dependence on ageing industries. Left-wing radicals such as M. Jean-Pierre Chevènement left the government at this point, although the two members of the Communist Party in the government soldiered on. Now the communists have departed, but M. Chevènement is back, as minister of education. What can be consistent about that?

It is important to remember the fine print. A year after his election, President Mitterrand startled the Western economic summit with his plea that the industrialized nations of the world should band together to use technology for the transformation of their economic prospects. By then M. Chevènement, originally minister of science and technology, had added industry to his portfolio and was attempting to carry through in France the policies that his president was urging on the West in general. There was never a chance that industrial innovation could prosper so quickly that France would hardly notice the disappearance of older uncompetitive industries, but Chevènement did his best. His successor, M. Laurent Fabius, has unfortunately been almost exclusively preoccupied with the downside of the last year's upheaval. But then and now, as Mitterrand's Prime Minister, Fabius has repeated the familiar doctrine that France will survive economically only by energetic application of technology through French industry. Can they succeed?

There is more than a slim chance. Even when the going has been roughest in the past three years, the government has sustained the scale of its support for basic science and for industrial development. The texture of research has been strengthened both in the universities and the research organizations, while there have been heartening technical successes such as the administration of the Ariane rocket programme and the operation of the huge French nuclear industry. The trouble of course, is that these are not yet (and may never be) economic miracles, while French ambitions to make a mark in the international market for telecommunications and computers remain as insubstantial as, say, the British. But plainly the new French Government does not intend to back away from this radical element of its economic policy. M. Fabius's post for science and technology is now filled by Professor Hubert Curien, one of the architects of the French space programme with his roots firmly in the academic side of the research enterprise — and with a high reputation elsewhere in Europe for his concern for international collaboration. Business as usual is the motto. □

Star wars

Pentagon's defensive criticism turned

Washington

THE latest acrimonious bout in the dispute between the US Department of Defense and the Office of Technology Assessment (OTA) seems to have ended in defeat for the generals. OTA has now firmly rejected a demand by Mr William H. Taft, Deputy Secretary of Defense, that it should withdraw its document published at the end of April which asserted that a star wars defence would be feasible.

The OTA study, described as a background paper and written by Dr Ashton Carter, now at the Center for Science, Technology and Foreign Affairs at Harvard University, has drawn the fire of senior Pentagon officials since its release by a Senate committee (see *Nature* 7 June, p.435). A list of technical criticisms was released to the press by Lieutenant-General James Abrahamson in May. Early the following month, Mr Taft wrote to OTA complaining that the document contained "serious technical errors" and that it misrepresented the basis of the administration's strategic calculations and demanding that it should be withdrawn.

This demand has now been firmly rejected by OTA, which has commissioned a *post hoc* review of the Carter paper from three people in a position to assess the technical data (many of them secret) on which the argument turns. In a letter to Mr Taft dated 13 July, Dr John H. Gibbons, the director of OTA, says that the review panel has said that the paper contains no technical errors or flawed assumptions that would seriously mislead readers, that it does not misrepresent the administration's goals (being largely silent on the subject) and that it is a "lucid, useful and generally reliable" introduction to the subject.

To the extent that much of the technical basis of the star wars defensive system is not publicly known, the argument between the Pentagon and OTA will be determined by the quality of the referees' opinions mustered by the two sides. OTA says that its referees were Lieutenant-General Glenn Kent, now with the Rand Corporation, Dr William Perry (formerly Under Secretary for Defense) and Professor Charles Townes (University of California, Berkeley). Mr Taft, in his letter of complaint, had said that "four independent organizations" had been critical of the background paper prepared for OTA.

The letter rejecting Taft's complaint does not deal with the technical issues in the dispute, described as an "episode of misunderstandings", but instead looks forward to "productive cooperation . . . as we strive together to serve the nation".

This is a reference to the more formal study of ballistic missile defensive systems which OTA, with the collaboration of the Pentagon, is to publish in segments during 1985.

In a memorandum to the Technology Assessment Board (the congressional oversight committee), however, Dr Gibbons is more outspoken about the reasons for OTA's confidence in its position. He says that while the Defense Department's "attack" on the OTA paper has "failed to make its technical case", general publication of the reasons for its rejection would "scarcely contribute" to obtaining "future assistance and cooperation" from the Pentagon.

Guessing at the reasons why the Pentagon's objections were off the mark, Gibbons says in his letter to his supervisory board that the complaints were assembled from comments by a number of defence contractors and "may not have been reviewed carefully". He also suggests that the accusation of "serious technical error", now rejected as invalid, may have arisen

because of an "effort to cast doubt on the judgements and observations which the Department of Defense finds unpalatable". He explains that most of the disagreement between OTA and the Pentagon (see below) arises because the latter has chosen to use different definitions in its calculations of the performance of a star wars defence.

This corrosive dispute between a government department and Congress's own technical advisory service faithfully reflects the division of opinion within the technical community about the feasibility of a nearly perfect defence against ballistic missile attack. Meanwhile, there has been a change of emphasis in what senior officials of the administration say about star wars.

President Reagan's statement on 23 March 1983, which offered the hope that "we could intercept and destroy strategic missiles before they reached our own soil or that of our allies", has been generally understood to refer to the destruction of all but a very small proportion of hostile missiles during their boost phase. But it is now clear that the administration's first goal is the development of a more effective anti-ballistic missile (ABM) defence for its strategic retaliatory force, with the prospect of destroying missiles in their boost phase something to toy with in the 1990s. □

Bones of contention

THE technical issues in dispute between the Pentagon and OTA are chiefly concerned with Dr Ashland Carter's calculations of the requirements of a system capable of destroying hostile missiles before the end of their boost phase at an altitude of not much greater than 200 km above their launching points. The following specific points arise:

- **Booster characteristics.** The OTA background paper argued that hostile missiles would be less easily detected (by the heat of their rocket motors) if they were designed so as to accelerate to their final velocity more quickly than is the case with liquid-fuelled rockets. (Carter points out that the MX missile now being built has some of these characteristics). The consequence is that the time available for detection by infrared sensors mounted on defensive satellites is reduced. The Pentagon's criticism of the OTA document says that it will be a decade before the Soviet Union could develop "fast-burn" boosters such as this, and that the result would then be a "70-90 per cent" reduction of the effective payload of ballistic missiles. In a letter to OTA, Major-General John C. Toomay, retired from the US Air Force, says that the complaint suggests that its authors "are confused" between fast-burn and "hardened" missiles (carrying protection against attack *en route*).

- **Battle stations.** The OTA paper made calculations of the number of satellites that would be needed to sense the emergence of

hostile missiles above the atmosphere of the Earth, and of the numbers of platforms carrying lasers or other "directed energy weapons" that would be needed to destroy them. The letter of complaint from Deputy Secretary Taft is reported to have alleged that the calculations were "wrong by at least a factor of two and, in some cases, a factor of fifty". Gibbons, in his memorandum to his board, says however that where OTA and the Pentagon deal with the same numbers of hostile missiles to be destroyed, the results differ by a mere ten per cent — 160 "battle stations" (OTA) rather than the 144 suggested by the Pentagon as sufficient.

- **Effectiveness.** Much of the disagreement between OTA and the Department of Defense turns on the assumptions made about the effectiveness of the various directed-energy systems being canvassed as components of future weapons systems. While the Pentagon's complaint says that OTA consistently underestimates the effectiveness, and the present state of development, of beam systems consisting of X-ray lasers and accelerators of neutral particles, General Toomay says in his letter of support to OTA that the "lethalities of various kill mechanisms [for destroying missiles] are essentially unknown at present", and that while a few years of research and development may clarify the potential of the beam weapons being considered, "we certainly cannot do this today". □

Nuclear waste

Ocean disposal recommended

Washington

THE US Government is being urged to consider ocean disposal for accumulating low-level radioactive wastes. In a special report to the President and Congress, the National Advisory Committee on Oceans and Atmosphere (NACOA) says it is concerned that "too little attention" is paid to the possible advantages of land disposal. The committee calls for more research to assess marine disposal options for ecological effects and pathways to man.

The United States has not dumped radioactive waste at sea since 1970, but other countries continued low-level disposal at various sites until a two year moratorium instituted under the London Dumping Convention in 1983. Despite strong political resistance to the practice, several nuclear nations are considering resuming dumping when the moratorium ends.

NACOA points out that the rate of production of low-level waste in the United States is likely to increase; the figure in 1980 was around 80,000 cubic metres. The US Navy has 100 operating nuclear submarines, and will be decommissioning between three and five per year, while several large commercial power reactors are nearing the end of their lives. NACOA believes that advances in marine science have made much less uncertain the environmental effects of ocean disposal, and urges Congress and the administration to reverse the present policy of excluding this option.

NACOA offers no detailed new study of environmental effects, but points out that all the anthropogenic radioactivity that has reached the oceans since 1944 is only one tenth of one per cent of the total natural radioactivity in the oceans. It raises the question whether agreement can be reached on a level of anthropogenic radioactivity low enough to be ignored. The difficulty, acknowledged by NACOA, is that the national phobia about nuclear activities makes rational debate difficult.

The federal government and the nuclear power industry have, it says, limited credibility — "too often power reactors have proved to be less safe than advertised, and proposed waste disposal sites have shown more complex geology than originally suggested". NACOA calls on the Department of Energy to provide financial support to citizens' groups to initiate a public debate on radioactive waste disposal and restore confidence.

The Office of Science and Technology Policy at the White House is charged with initiating the new research effort, aimed at clarifying the effects of low-level waste disposal at sea. At present, responsibilities are divided between several agencies, and NACOA wants to see a new inter-agency forum to examine the issues in detail. Investigations under way into the feasibility of sub-seabed emplacement of high-level waste might, it believes, shed light on low-level waste options. It also calls for an inventory of all anthropogenic radioactivity in the ocean; the International Atomic Energy Agency could undertake the task. In addition, US waste classification should conform with that of the agency.

According to NACOA's own estimates, the largest source of anthropogenic radioactivity in the ocean has been from weapons tests before 1968. Other major contributors have been lost submarines, plutonium-producing reactors and reprocessing plant. The list, however, is undoubtedly incomplete. NACOA bemoans the fact that much US dumping of low-level waste before 1970 was not adequately recorded, so that the opportunity to study carefully the effects of known dumps has been lost.

International agreement on dumping is highly desirable, it says, because the United States is not the only country contemplating a return to sea disposal, and its actions will be seen as setting a precedent.

Tim Beardsley

Burford in hot water

FEELINGS continue to run high over President Reagan's appointment of Mrs Anne M. Burford as chairman of NACOA, the National Advisory Committee on Oceans and Atmosphere (see *Nature* 19 July, p.171). On 24 July, the Senate passed a resolution calling on the President to withdraw Mrs Burford's nomination by 74 votes to 19, with strong support from both Democrats and a majority of Republicans.

The motion was sponsored by Senator Edward Kennedy, who told the Senate that the President appeared to have forgotten Mrs Burford's record as administrator of the Environmental Protection Agency, described as "controversial and flawed". Senator Kennedy proceeded to recite a catalogue of delays and obstructions to the agency's work which he laid at Mrs Burford's door. Senator Alan Cranston,



co-sponsoring the motion, described Mrs Burford as an "enemy of environmental protection". The Senate's confirmation is not required for Mrs Burford to be appointed, because the committee on oceans and atmosphere is only advisory. President Reagan appears unrepentant, however. Noting that the Senate's resolution was not binding, the President has said he will stand by the appointment.

The matter is nevertheless unlikely to rest there. Moves are afoot in both houses of Congress to change the status and powers of NACOA. A bill introduced last week by Senator Ernest Hollings would at a stroke abolish the committee and then reconstitute it, but in such a way that members would have to claim some scientific expertise relevant to environmental policy (which it is thought would rule out Mrs Burford).

Under the Hollings bill, the Senate would have the power to veto appointments. There may be some constitutional difficulties in this course, however, and an alternative being promoted in the House of Representatives by Representative Richard Ottinger would do away with NACOA altogether. This might pressure the Senate into reviving an earlier proposal for a National Oceans Policy Committee, which again would increase the power of Congress to influence appointments.

Tim Beardsley

Uranium

Australian exports continue

Canberra

FOLLOWING a 55:44 vote in favour at last month's Labor Party national conference, the continued export of uranium oxide is now part of Australian government policy. The most important in a week-long series of victories for Prime Minister Mr R.J. (Bob) Hawke, the vote was carried amid impassioned pleas and protests from the left, newly fortified by recent attacks on Mr Hawke's major technical prop, the Slatyer report (see *Nature* 309, 647; 1984), and the appearance of other commentaries condemning the export of yellowcake, even under strict bilateral and multilateral safeguards.

The new policy means that the two existing Northern Territory mines, Ranger and Nabarlek, may negotiate new export contracts and that development may begin on the giant Roxby Downs ore body in South Australia.

If French underground nuclear tests in the South Pacific do not end soon, the policy also carries with it a commitment towards implementing a halt to uranium shipments to France, the next of which is due in October. The uranium export industry will be worth an estimated \$A360 million this year, rising, according to the Slatyer report, to as much as \$A1,000 million by 1993.

Jeffrey Sellar

US immigration

Reprieve by default for academics

Washington

IMMIGRATION legislation that could restrict university recruitment seemed to be foundering last week in a storm of racial sentiment and pre-election politicking. The controversial Simpson-Mazzoli bill is proving to be a sound stick with which newly-enthused Democrats in Congress can beat the Reagan Administration. The possible effects on scientific manpower in the United States seem to be largely overlooked.

While national attention focuses on the bill's provisions to apply an amnesty to illegal immigrants, universities seem to have resigned themselves to a back seat. The bill originally required foreign nationals studying in the United States to return to their home countries for a period of two years after graduating, before being allowed back into the United States. Not surprisingly, many of those who recruit graduates — both inside and outside the universities — were dismayed, and both houses of Congress accepted that this blunt instrument had to be sharpened.

In its present form the bill will allow exemptions to the "two years non-residency" requirement, but the Senate and House of Representatives versions still differ. Educational lobbyists seem to have had their energies drained by their earlier efforts, and say they cannot expect to have any more influence on the bill. The exemptions will apply only to those who have been offered employment in the United States in the same field as their degree; others will be denied visa extensions. If the employment offered is on a faculty, the subject restrictions are waived; otherwise exemptions to the two years' foreign residency requirement are to be limited to those with degrees in natural science, mathematics, computer science or engineering.

The position of the American Council on Education, an umbrella body for various groups, is that the *status quo* should be preserved, with no new restrictions on recruitment. The bill will, if passed, at the very least make extra work for university administrators and could yet hinder recruitment. One of the difficulties admitted by the council is the lack of data on foreign nationals qualifying and working in the United States, although according to one survey conducted by the National Academy of Sciences something like 55 per cent of foreign nationals who obtain a PhD in the United States remain for long enough to seek a visa extension. But there is nothing like a complete profile of foreign students' employment, which is in itself hindering efforts by those representing the interests of universities and high-tech industry.

Engineering groups have been particularly concerned about the bill in the

past, but, like the universities, now seem resigned to whatever Congress hands down. The issue of labour certification — the need to demonstrate that no US citizen can undertake a job awarded to a foreign national — is one area where the new bill could overturn the conventions that have been established. In particular, it is now to be questioned whether the outcome of an academic search will automatically satisfy the Department of Labor that a work visa is justified.

UK fast breeder

Accountants' worries

THE lack of major specific targets in the United Kingdom's development programme for commercial fast-breeder reactors has led to weakened financial control, according to a recent report of the influential House of Commons Public Accounts Committee*. The report urges that the Department of Energy and the UK Atomic Energy Authority (UKAEA) should be attempting to specify intermediate milestones — in terms of both achievement and cost — to enable the success of the programme and its cost effectiveness to be more easily assessed.

Following the government's decision to slow down fast-breeder development in 1982, UKAEA is cutting its annual expenditure on the project by a third to about £70 million a year. According to the authority's chairman, Sir Peter Hirsch, this is the minimum commensurate with effective participation in the European collaboration on fast breeder design (see *Nature* 307, 200; 1984). In his evidence to the committee, he reported that the total cost of a development programme leading

The House of Representatives passed its own version of the bill by the narrowest of margins a few weeks ago, but is unlikely to allow any further concessions to the stricter Senate version. Senator Alan Simpson is still insisting that the bill will survive, although hopes of appointing a conference committee to seek a truce with the House before the forthcoming Republican convention have now been abandoned, apparently to spare the bill's supporters embarrassing political pressure. Others are by now convinced that the bill will be a hostage to the presidential election and will soon disappear without trace.

Tim Beardsley

to commercial power stations was estimated to be £3,100–£3,700 million, depending on assumptions about future interest rates. This calculation assumes that a Commercial Demonstration Fast Breeder Reactor (CDFR) is constructed in 1993 and that power stations come on-line in 2005.

The Public Accounts Committee acknowledges the uncertainties in such calculations but notes that this is the first time that such a figure has been presented to Parliament. And because there is no government commitment to the construction of a CDFR, UKAEA's expenditure is set at an annual rate that cannot be judged by a specific timetable of development. Nevertheless, the committee urges that intermediate targets should, where possible, be agreed upon with the Department of Energy and costed accordingly.

Philip Campbell

*Thirty-second Report from the Committee of Public Accounts, Session 1983-84, *Development of Nuclear Power* (House of Commons Paper 367, HMSO, £4.15).

COCOM agreement on computers

Washington

COMPUTER manufacturers in the United States seem on balance to be relieved by the new agreement on computer export controls reached last month by COCOM, the international co-ordinating committee on strategic exports. The new rules will put US and European manufacturers on a more equal footing. But while some control limits were raised, new restrictions were imposed for small personal computers and software.

The agreement is seen as a major breakthrough in the vexed international issue of export controls: US attempts to enforce extraterritorial provisions in its own export control laws have been a continuing source of friction. The 15 member nations of COCOM (members of the North Atlantic Treaty Organization, minus Spain and Iceland, plus Japan) have been discussing the

question for almost two years.

The maximum "processing data rate" of exported mainframe computers has been increased from 32 million bits per second to 48 million bits per second. For personal computers, the cut-off point will be between 7 and 8 million bits per second: the United States had been particularly keen to see controls on the smaller machines, which can have battlefield applications.

The new software controls will, it is believed, apply only to sophisticated real-time functions and will not cover standard programs. Telecommunications and switching equipment are also covered. But, despite the relief that agreement has been reached, there is still some frustration among the manufacturers that the details are to remain confidential, although some guidelines will probably be published.

Tim Beardsley

US election

From frying-pan to Senate?

Washington

SINCE the withdrawal earlier this year of Senator John Glenn from the competition for the Democratic presidential nomination, interest in the electoral fortunes of technically-minded members of the US Congress has centred on Representative Albert Gore's attempt to win the Senate seat being vacated by Senator Howard Baker, now the Republican leader in the Senate. Although Gore is a Democrat, his reputation in Washington, the strength of his roots in Tennessee and the swing of the political pendulum are reckoned to conspire to his advantage in November's general election.

During his eight years in Congress, Gore has principally made his mark as chairman of the investigations and oversight subcommittee of the House of Representatives Committee on Science and Technology. By using the subcommittee as a platform, from which he has launched inquiries into



Representative Albert Gore

topics such as scientific fraud and the effects of automation on employment, Gore has provided yet another proof that virtually any power base will serve the needs of one who is sufficiently resourceful and energetic.

At 38, Gore is as energetic as they come. His staff boasts of his voting record, saying that he has been on hand for 78 per cent of roll-calls in the House of Representatives during the past eight years. Philosophically, his line on science and technology is a blend of optimism and anxiety, that of an upbeat Alvin Toffler.

Just now, Gore is especially exercised by the question whether the coming of the fifth generation of computers (and "then the sixth generation") will cause intolerable social strains. While accepting that increased productivity may create extra jobs for those displaced from manufacturing, he is concerned that there may be insufficient opportunity for retraining and wonders "whether we will be imaginative enough to devise ways of compensating adequately" those destined to take jobs now badly paid.

John Maddox

Laboratory animals

Berkeley fined for violations

Washington

THE University of California at Berkeley has been fined \$12,000 and ordered to correct long-standing deficiencies in its treatment of laboratory animals, the US Department of Agriculture (USDA) said last week.

Under the "consent decision" reached with the university, there was no admission of wrongdoing, but the university agreed to pay the fine — believed to be the largest ever assessed against a university in such a case — and take several specific steps to assure compliance with the requirements of the federal Animal Welfare Act. The university will have to run a training programme for those involved in animal care and will be required to establish an animal care committee that will oversee Berkeley's efforts to comply with the federal standards. At least one member of the committee must be a member of the community, not affiliated with the university, and involved neither in research nor the animal welfare movement. The committee must make quarterly reports to the chancellor of the university and to USDA for at least three years.

The university also moved last year to allocate half a million dollars for major renovations in its existing laboratory facilities. Two additional campus veterinarians were hired, a training programme was established and — apparently for the first time — the chief veterinarian was assured access to all animal facilities on campus.

The USDA fine was the culmination of years of efforts to get the university to correct apparent violations of the Animal Welfare Act. USDA inspection reports cite uncleaned cages, overheated and poorly ventilated rooms, overcrowding of animals, and feeders contaminated with urine and faeces. Numerous inspection reports note inadequate veterinary care or a lack of trained employees.

An attorney for USDA said the problems at Berkeley were primarily the result of careless attitudes and a failure of organization. He said there was no one with undisputed authority to order repairs or to have poor practices corrected.

Berkeley will be permitted to apply \$10,000 of the \$12,000 fine to its training programme.

Stephen Budiansky

Koshland for Science post

Washington

THE new Editor of *Science* magazine will be Dr Daniel E. Koshland Jr, professor of biochemistry at the University of California, Berkeley. Professor Koshland's appointment was announced last week by David Hamburg, president of the American Association for the Advancement of Science, and will take effect from 1 January 1985.

Koshland replaces Dr Philip Abelson, who has served as Editor of *Science* for the past 22 years. He will divide his time equally between the editorship and his university teaching and research in California, while a deputy editor (who is yet to be appointed) will look after much of the day-to-day running of the journal.

Koshland received his BSc at Berkeley in 1941 and his PhD from the University of Chicago in 1949. His research areas have been the regulation and control of enzyme activity, and more recently information processing in sensory cells. At present he is using recombinant DNA techniques to analyse components of chemotaxis in bacterial model system.

He is, he says, inheriting an "outstanding" journal from Dr Abelson, but denies that he will be a caretaker editor. He expects to be making some changes to *Science* to reflect the changing times and his own interests and says it will be a challenge to continue the progress established by his predecessor. While declining to be drawn on specifics, he does not rule

out moves to increase the target group of the publication.

Koshland will relinquish his other editorial positions to devote all his time outside university to *Science* and will also be reducing his university teaching load, although the details of how he will divide his time between two jobs two and a half



Daniel Koshland

thousand miles apart have yet to be settled. He expects to be active in soliciting leading articles and evaluating research reports. Does he see *Nature* as a major competitor? Possibly. But, he says, "there's room for both of us".

Tim Beardsley

Japanese nuclear power

Lawsuit dismissed

Tokyo

A LAWSUIT aimed at forcing the Japanese Government to withdraw permission for the construction of a nuclear power reactor in Fukushima prefecture, some 250 km north of Tokyo, has been dismissed — nine and a half years after it was filed by a group of 400 local citizens.

The decision, the first in Japan since the Three-Mile Island incident in the United States in 1979, will come as a boost for the nuclear power industry which, in marked contrast to that of the United States, is still booming and self-confident. In the week that saw the end of the court hearing, power companies announced ambitious plans for construction of a complete nuclear fuel cycle facility — uranium enrichment, reprocessing, low-level waste storage and probably high-level waste vitrification — in Aomori prefecture, a government advisory council recommended doubling the number of nuclear power plants by 1993 and the government began talks in Beijing to finalize a bilateral agreement between Japan and China on nuclear power utilization.

The citizens' suit had centred on claims that permission for construction had been given "illegally" and "irrationally". But most of the court's time — some six years — was spent not on these issues but in deciding whether the citizens had any right to bring the matter to court at all. On this first point the plaintiffs were victorious: the defendants' argument that the nuclear reactor control law refers only to the "public interest in general" was dismissed and it was accepted that those exposed to the possibility of suffering from a nuclear power plant accident have a right to protection of their personal interests. A precedent has thus been set for other lawsuits, including four now in progress.

Afghan students

Budapest

AFGHAN students in Hungary are extremely concerned about their future. During the past month, the Afghan Embassy in Budapest has called in their passports, offering no explanation. The students fear that this may be a preliminary to their being called home for military service without completing their studies.

There are believed to be several hundred Afghan students in Eastern Europe, and even more in the Soviet Union, mainly studying scientific and technological subjects. There is no information so far as to whether the call-in of passports applies only to Hungary or whether it has also been taking place in the "harder-line" countries such as Czechoslovakia and East Germany.

Vera Rich

The court also took account of the plaintiffs' questioning of the legal powers of the government to authorize nuclear power construction. What was at issue was whether the government can use "discretionary powers" to authorize the construction of a plant which may threaten the lives of a large number of people. The court recognized the government's rights but at the same time ruled that the seriousness of a nuclear accident meant that the scope of that discretion is very small — the government must be able to prove that the judgement about the safety of a reactor had been properly made.

The plaintiffs were disappointed on what they saw as the real issue of obtaining a ruling on whether the reactor was safe: doubts had been raised about the safety of the fuel rods, the emergency core cooling system, the handling of nuclear waste, the impact of thermal effluent and the possibility that local residents would be continuously exposed to low-level radiation even during the normal operation of the reactor. But despite — or perhaps because of — the many tons of documents provided by the defendants, the court settled the matter on purely legal technicalities and failed to provide any broad judgement on nuclear reactor safety. The court was able to declare only that there had been no violation of the legal procedures in the process of safety screening and that the government's decision had not been made "irrationally". In so doing, the court accepted that the scope of the government inspection was restricted to the reactor itself — that is, to an examination of the basic design. All worries about the safety of waste and operation are thus left to post-construction regulations. Evidence from the Three-Mile Island incident was also considered inapplicable.

That judgement would eventually go the government's way has never really seemed in very much doubt — the reactor has after all already been built and is in operation. But the plaintiffs still feel that the basic question of who is going to assure the safety of the reactor has not been answered. For Fukushima residents it is not just this particular plant that is worrying: with eight nuclear reactors on two sites in the area, two more reactors and a prototype heavy-water plant under construction, they are surrounded by the world's highest concentration of nuclear plant.

The citizens' group is now to appeal to a higher court. Most commentators feel, however, that a court of law will never be able to give them the kind of assurance they seek. Instead there have been renewed calls for moves to strengthen the system of public hearings to help build public confidence in nuclear reactor safety.

Alun Anderson

Operation of power plants

Tokyo

JAPAN has 25 nuclear power stations in operation, which puts it in fourth place after the United States (82), France (36) and the Soviet Union (35). In contrast to the United States where no new contracts have been awarded in the past five years, Japan is continuing an ambitious programme of nuclear power plant construction.

Ten plants are being built and research is being hurried forward on advanced boiling-water reactors that use a concrete rather than a steel containment vessel (in cooperation with US builders), on heavy water reactors and, with so far unshaken optimism, on fast-breeder reactors.

Plants now in use — almost all boiling-water and pressurized-water reactors based on General Electric and Westinghouse designs — produce 17 per cent of the nation's power. With average operating efficiency close to 70 per cent, the plants approach perfect running: they must be shut down at least 25 per cent of the time for safety checks. Japan's nuclear safety record also appears remarkably good — no injury or loss of life has been caused by radiation leaks — and plant operators boast that the Three-Mile Island accident could not have happened in Japan where operator training standards are higher. Government documents show, however, that around 25 accidents and breakdowns are reported each year.

At least one of these accidents, which involved the drainage of dangerous material into the sea, was potentially serious and led to the disclosure of other mishaps that had been kept secret.

Since 1978, public hearings have become a part of the licensing system. In a first hearing, the power plant builder presents its plans and tries to win local support; after reports from the builder and the Ministry of International Trade and Industry (MITI), the Nuclear Safety Commission then provides experts to answer questions at a second hearing (albeit with some limitations on follow-up questions). The final decision on construction is then made by MITI council. As MITI is extremely active in promoting nuclear power development, it is difficult to see how it can act simultaneously as a completely neutral licensing body. Perhaps partly because there is a feeling that any ruling is a foregone conclusion, public hearings have so far attracted more demonstrators than participants — the latest one, held in June in Saga prefecture, required the presence of 1,800 riot police. But to the power plant builders, this tendency simply demonstrates the lack of any real rational grounds for opposition to nuclear development.

Alun Anderson

Price of European journals

SIR — Over the past three or four years, we have become increasingly concerned because some European, particularly British, publishers charge higher prices to North American readers than to those in other parts of the world.

One may well ask why *Nature* costs \$230, or about £160, in the United States, whereas it costs only £98 in the United Kingdom and £120 in the rest of the world. The surprise and vexation of American librarians becomes greater when we have to pay \$588 (£405) for the *Philosophical Magazine*, which costs only £207 elsewhere.

We could expand this list. Unfortunately, the practice is spreading among both commercial and society publishers. But the two examples suffice to illustrate the problem. One shows a questionable but still modest surcharge of 30 per cent, the other a not unusual doubling of the price for US subscribers.

When asked to explain this practice, publishers tend to cite the cost of air freight and unstable exchange rates. But air despatch is neither limited to North America nor more costly than it is to other overseas countries. Nor can this cost justify the difference in prices. Foreign exchange rates have moved for several years in one general direction and have consistently benefited the foreign publisher who early sets his price in dollars, not to speak of the one who sets it at unrealistic levels. When, years ago, the dollar was steadily declining in relative value, the same publishers did not hesitate to issue supplementary invoices or to demand payment in their countries' currency. What seemed fair then, is fair now. If publishers demanded payment in pounds sterling then, they cannot in good faith demand it in dollars now.

The American representative of a major British university press recently let the cat out of the bag by noting that the higher prices were not based on exchange rates or shipping costs but were the result of the opportunities presented by "different markets".

We find these practices unacceptable. They put a further library cost burden of perhaps several million dollars on researchers in the United States and only aggravate an already critical situation.

Apart from direct but futile correspondence with publishers and subscription agents, we have brought our objections to the attention of the American Library Association, the Special Libraries Association and the Association of Research Libraries. One of us is preparing a paper on the topic for publication in a US library journal, in which he intends to use and expand on the material in this letter.

We have cancelled some British journals and have not subscribed to some new ones because of higher prices for US libraries.

For some journals from other European countries we have been able to take other measures, and we are exploring the possibility of similar arrangements in England.

LARRY X. BESANT
SIEGFRIED RUSCHIN

*Linda Hall Library,
5109 Cherry Street,
Kansas City, Missouri 64110, USA*

● Until October 1975, the price of *Nature* in the UK was £27.50, while US subscribers were supplied by surface mail (at £31 a year) or by airmail (£50 a year). Later in 1975, in October, arrangements were made to supply US subscribers by air-speed delivery and the price of *Nature* was fixed at £35 in the United Kingdom and \$95 in the United States. I am sure that with the dollar converting at more than two to the pound and with the extra cost of air-speed, the prices were virtually identical. Since the end of 1975, the price of *Nature* has increased from £35 to £98, a ratio of 2.7, while the price of a full subscription in the United States has been increased to \$230, a ratio of 2.4. Since November 1982, the price of *Nature* in both countries has remained essentially unchanged, and will not change until at least the end of 1985. But unhappily, with the rapid decline of sterling relative to the dollar, the US price seems relatively much greater than that in the United Kingdom. But it is worth also saying that since we now manufacture American copies in the United States and are thereby (we hope) able to provide an even faster service than air-speed allows, we have to pay substantial printing bills in the United States and are thus less able to profit from the dollar's strength. — Editor, *Nature*. □

Tropical malaise

SIR — The current decline in tropical ecology pointed out by Norman Cole (*Nature* 309, 204; 1984) is one for which there is no immediate prospect of reversal. The problem is one of the availability of appropriately trained manpower and extends into applied areas such as agriculture and fisheries. There is a significant demand for people of this type from developing countries in the tropics by the international agencies and overseas universities but the rate of replacement appears much slower. Often the most promising individuals in a government department of a developing country disappear abroad, understandably to take up more rewarding appointments in these situations, with no qualified junior who might profit from the increased experience to take over. There appears, therefore, to be a net loss of people at the highest level which exceeds the current availability of trained experienced people.

In the developed countries it is not just

that economic stringency has led to a reduction in the number of projects in the tropics but, more to the point, the number of opportunities for recent graduates to gain tropical experience is becoming severely limited. This is particularly noticeable in a country such as Britain which has a substantial tradition of work in tropical biology; however, the number of outlets for younger scientists to gain the crucial initial experience are now very few, and are principally through the voluntary agencies, with a few through government technical assistance programmes. This can only lead within the next decade or so to a much reduced commitment by this country to the tropical sciences. A few countries, such as Norway and the Netherlands, are increasing the opportunities for their scientists, including those newly qualified, to work in the tropics, which will provide a cadre of experienced workers in the medium term, although the total numbers may not be so great.

All of these factors add up to a rather discouraging prospect. The sparse distribution of experienced scientists from the developing countries and the international drain on these resources in conjunction with the progressive limits on the possibilities for young scientists from the developed countries to gain the all-important first experience in the tropics could well precipitate a manpower crisis in ecology and food production sciences in ten to fifteen years' time.

A. I. PAYNE

*Department of Biological Sciences,
Coventry (Lanchester) Polytechnic,
Coventry CV1 5FB, UK*

Sea of subdued light

SIR — I wonder whether there is any interest in Ruskin's account of Homer's sea? Ruskin says¹:

As far as I can trace the colour perception of the Greeks, I find it all founded primarily on the degree of connection between colour and light; the most important fact to them in the colour of red being its connection with fire and sunshine; so that "purple" is, in its original sense, "fire-colour", and the scarlet, or orange, of dawn, more than any other fire colour. I was long puzzled by Homer's calling the sea purple; and misled into thinking he meant the colour of cloud shadows on green sea; whereas he really means the gleaming blaze of the waves under wide (sic) light. Aristotle's idea (partly true) is that light, subdued by blackness, becomes red; and blackness, heated or lighted, also becomes red. Thus, a colour may be called purple because it is light subdued (and so death is called "purple" or "shadowy" death); or else it may be called purple as being shade kindled with fire, and thus said of the lighted sea; . . .

J. BENTLEY

*179 Howlands,
Welwyn Garden City,
Herts AL7 4RL, UK*

¹ Ruskin, J. *The Queen of the Air: Being a Study of the Greek Myths of Cloud and Storm*, 110-111 (George Allen, Orpington, 1869).

Verification limits for test-ban treaty

from Lewis A. Glenn

In spite of improvements in seismic techniques, the remote identification of seismic events as nuclear explosions remains limited.

BANS or limitations on nuclear weapons tests are one of the many types of arms control measures that have been attempted over the years. The first nuclear test agreement, the test moratorium, was made in 1958 and lasted until the Soviet Union unilaterally resumed testing in the atmosphere in 1961. It was followed by the Limited Test Ban Treaty of 1963, which prohibited nuclear tests in the atmosphere, in outer space and underwater. In 1974 the Threshold Test Ban Treaty (TTBT) was signed. This treaty limited underground tests after March 1976, to a maximum yield of 150 kilotons. (1 kt is the energy release equivalent to 1,000 metric tonnes of TNT.) The TTBT was followed by a treaty limiting peaceful nuclear explosions and, although neither of these was ever ratified by the US Senate, both the United States and the Soviet Union claim to be abiding by the 150 kt yield limit. A comprehensive test ban treaty (CTBT), prohibiting all testing of nuclear weapons, has also been discussed and there is widespread public interest in this idea. However, a verifiable CTBT is a contradiction in terms. No monitoring technology can offer absolute assurance that very-low-yield illicit explosions have not occurred.

The concept of *adequate verification* has been employed in an attempt to avoid this pitfall. This term is usually understood to mean "that which would reduce to an acceptable level the risk that clandestine test programmes of military significance could be conducted..."¹. However, it has proved difficult to obtain agreement on what is meant by an acceptable level of risk and, especially, programmes of military significance.

An alternative approach is to characterize the limitations of realistic monitoring and analysis methods and the potential for clandestine testing. A detailed study along these lines has been prepared by J. Hannon⁴ of our laboratory and is summarized below, along with a brief but comprehensive review of cavity decoupling, the most serious evasion threat. When these constraints and opportunities are taken into account, the outlines of a max-

imally restrictive, yet verifiable, test ban agreement can be brought into sharp focus.

Tools for verification

The principal tools for monitoring compliance with a CTBT are seismic networks and surveillance satellites. On-site inspections might also be required to resolve ambiguous events. Satellites are actually of limited use, since it is possible to carry out low-yield explosions in buried cavities without any visible ground-surface motion resulting. Moreover, test activities, including the required excavations, can be associated with mining or other large-scale industrial undertakings. The critical element of the monitoring system is thus the network of seismic stations, and in particular the in-country stations.

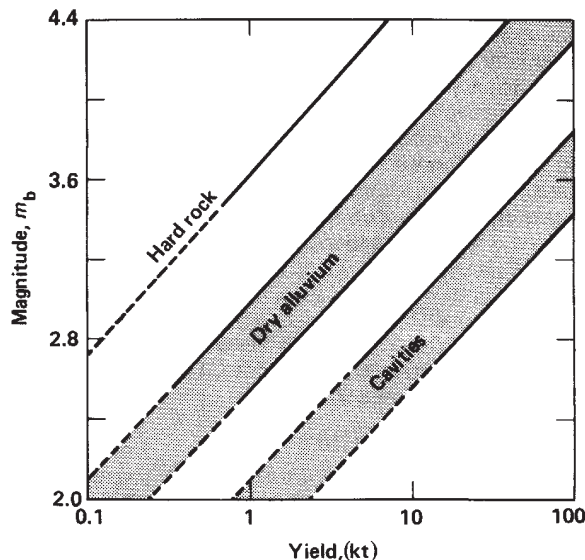
Internal stations provide much more useful data than do stations situated outside the borders of testing nations. The external stations are generally so far from potential evasion sites that they record the motion of only a few waves from the larger events (those with seismic magnitudes, m_b , greater than about 4 on the Richter scale) and high frequencies have been significantly reduced by attenuation and scattering throughout the Earth. The larger amplitudes recorded by the internal stations, for the same event, result in an increase in detection capability. The multiple waves and

higher frequencies improve the ability to distinguish between explosions and earthquakes. These two functions, detection and identification, are the main elements in the verification process.

Detection capability

At any given point on the surface of the Earth, the detection capability of a seismic network is specified in terms of the seismic magnitude of the source that will be detected there with a specified degree of confidence. Algorithms have been devised by independent investigators to estimate this capability as a function of various input parameters^{4,13}. The predictions are in remarkable agreement as long as the same ground rules are observed. Figure 2 shows the estimated overall detection capability as a function of the number and type of internal stations; the calculations were made by J. Hannon of our laboratory⁴. In this figure, the overall detection capability is defined as the magnitude above which an explosion at any point in 90 per cent of the Soviet Union would be detected with 90 per cent confidence. The capabilities of simple internal stations (consisting of single, three-component seismometers) are contrasted with those using regional arrays, in which 16–25 seismometers are deployed over an area of about 10–100 km². The array stations have a lower effective noise level and also have improved ability to record

Fig. 1 Seismic body-wave magnitude against explosive yield for different explosion environments. The relationships are based on US data from the Nevada Test Site.



Lewis A. Glenn is Associate H-Division Leader at the Physics Department, Lawrence Livermore National Laboratory, University of California, PO Box 808, Livermore, California 94550, USA.

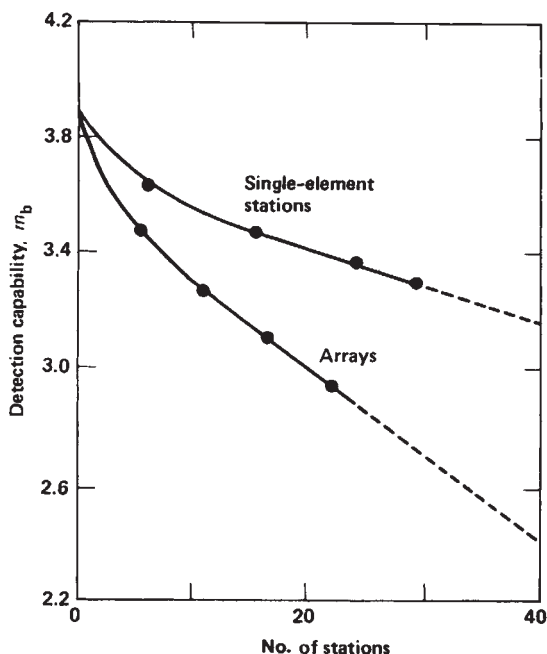


Fig. 2 Overall detection capability as a function of the number and type of internal stations. The overall detection capability is the magnitude above which an explosion at any point in 90 per cent of the Soviet Union would be detected with 90 per cent confidence.

combinations of regional waves.

Figure 3 shows the estimated effects of varying some of the parameters implicit in the calculations of Fig. 2. P_g and L_g are large-amplitude seismic waves propagating in the crust and upper mantle and recorded at regional distances; they are a primary reason for the increased capability of regional stations.

The two parameters that have the most significant impact on the detection capability are the SNR and the required degree of confidence, the second of which accounts for the main differences in detection capability reported by different groups^{4,13}. Some believe that any testing by the other side under a CTBT could lead to a significant threat to national security and that the probability is high that evasion will be attempted irrespective of the potential political costs. They insist that the degree of confidence in the ability to detect (and identify) a single clandestine test should approach 90 per cent. Others have argued that only a programme of repeated testing has military significance and contend that the anticipated costs of being caught are so high that even a small possibility of detection is a significant deterrent to any potential evader. Those holding this view suggest that any verification provisions need have a degree of confidence of only 30 per cent or less. One way of comparing the two approaches is to note that, given a 30 per cent degree of confidence for detecting a single event as an explosion, a series of seven tests would have to be carried out before the probability of detecting at least one CTBT violation would exceed 90 per cent.

Seismic identification

Once seismic waves from an event have been detected, the source must be identified as a nuclear explosion, a chemical explosion or an earthquake. Many

discriminants have been proposed. First, the located depth of the source is useful, since current drilling limits are less than about 10 km. Therefore, explosions are not likely to be responsible for signals originating at greater depths.

Second, the areas in which large magnitude earthquakes occur are fairly well known. A strong seismic signal originating at a shallow depth in, say, Kansas is much more likely to be the result of an explosion than an earthquake; the same would not necessarily be true for California. However, areas that are seismically inactive at high magnitude levels may not be so at the low levels demanded for CTBT verification. For example, a recent study of parts of Louisiana, Oklahoma and Texas, which had shown only one earthquake of magnitude 4.5 or greater in 8 years, revealed that there is almost one earthquake per day in the magnitude range 2.0 to 3.9 (ref. 14). Similar results can be expected in many granitic regions, so that the areal discriminant is probably of limited use for small events.

For those large events that are not eliminated by depth or location, one of the most useful discriminants is based on the ratio of surface-wave to body-wave magnitudes ($M_s : m_b$). If an explosion and an earthquake have the same body-wave magnitude, the surface-wave magnitude for the earthquake is generally larger^{15,16}. It has yet to be shown, however, that $M_s : m_b$ is useful at low magnitudes, especially when explosions are set off in long tunnels or odd-shaped cavities (see below).

A number of other promising regional discriminants have been suggested, based on differences in the spectral content and radiation pattern of the recorded signal¹⁹, but their use requires a greater SNR than for detection alone. It is generally accepted that almost all events having a SNR three to

four times (0.5–0.6 mag or 10–12 dB) greater than that required for detection will be identified. This implies that some events will be detected but not identified, thereby generating “false alarms” that would have to be resolved by on-site inspections or other means. A rough estimate of the number of such events can be derived from Fig. 4, which plots the average number of shallow earthquakes above a given seismic magnitude; small events, with $m_b < 3.5$, are poorly recorded, so the data are extrapolated below this level. An optimistic assumption is that 80 per cent of the events in the magnitude interval between 2.7 and 3.2 (the level of detection capability reached by a network of 25–30 regional arrays) will be identified. This leaves several hundred unidentified events a year, which would pose an obstacle to continuing confidence that the terms of a treaty were being observed.

Opportunities for evasion

Detection of an explosion can be avoided by ensuring that the signal generated is below the background noise level, by emitting “normal” signals at times of high noise, or both. If the noise only partially masks the signal, detection may occur but identification may be impaired. Reducing the signal strength at the receiving station may be accomplished either by reducing the coupling of the explosive energy into the Earth at the source or by choosing (through site selection) a path from the source to the seismic station through a region that absorbs significant seismic energy.

To understand the significance of decoupling, it is necessary to interpret the detection capability of a network in terms of explosive yield, as illustrated in Fig. 1. The seismic body-wave magnitude is plotted against the yield in kilotons. The data are representative of US experience at the Nevada Test Site². Their applicability to the Soviet Union is uncertain. Various investigators^{18,19} have presented evidence that a given explosion in the Soviet Union registers several tenths of a magnitude unit higher than the same explosion at the Nevada Test Site (which is a principal factor in the recent controversy over alleged violations of the Threshold Test Bank Treaty).

The upper line in Fig. 1 shows the seismic magnitude that would be generated by explosions that are tamped, or well-coupled to hard rock. Under these conditions, if the lower limit of detectability (from Fig. 2) corresponds to $m_b = 2.7$, the maximum explosive yield would be approximately 0.1 kt. It should be noted that, although the domain in Fig. 1 extends to this low yield, there are very few hard-rock data below 1 kt; these are characterized by considerable variability and there are also problems in obtaining a consistent definition of magnitude. The line drawn in Fig. 1 is based on the “granite” representation in ref. 2; both lower³ and higher⁴ estimates have

been made for hard rock.

The middle band in Fig. 4 shows the effect of detonating the explosion in the deep, dry alluvial deposits at the Nevada Test Site. For the same body-magnitude, the explosive yield would be about 10 times greater in alluvium than in hard rock. The reason is that the alluvium is porous, so that a large fraction of the source-energy is dissipated by pore collapse. It is uncertain whether there are similar geological formations in the Soviet Union.

The lower band in Fig. 1 corresponds to explosions in large cavities and represents by far the most important potential for evasion.

Cavity decoupling

The idea of using a large cavity to muffle the seismic signal from an underground explosion was first proposed 25 years ago by Albert Latter and Hans Bethe at an early conference on the discontinuance of nuclear tests⁵. They calculated that the strength of the distant seismic signal in an elastic medium is determined primarily by the volume of the cavity created by the explosion. If the explosion is set off in a large cavity, then it is the differential increase that determines the signal amplitude. Moreover, if the cavity is just large enough for its walls to suffer only elastic deformation, the distant signal, for a given yield, is minimized and it does not pay to make the cavity any bigger. Inelastic behaviour causes increased coupling because the medium is unable to support shear and can thereby flow and thus undergo large displacements.

Experimental confirmation of the cavity-decoupling theory was first obtained in a series of experiments, collectively called COWBOY, conducted with chemical explosives in a Louisiana salt mine in 1960⁶. Salt was chosen as the test medium because of its homogeneity *in situ* and because of the ease in which a cavity can be constructed and maintained. The ratio of the seismic amplitude in the tamped explosion to that obtained when the same yield is generated in an excavated cavity is called the decoupling factor; a decoupling factor of 100 was measured in COWBOY.

Confirmation of cavity decoupling in a nuclear experiment was made in the STERLING event (0.38-kt yield)⁷, which was fired in the Tatum salt dome (near Hattiesburg, Mississippi) on 3 December 1966, more than two years after the SALMON event (5.3 kt yield)⁸ created the cavity. A maximum decoupling factor of 70 ± 20 was determined for STERLING, indicating that full decoupling may not have been achieved. It was known before STERLING that the SALMON cavity (17.4 m radius) was too small fully to decouple the explosion, which would have required roughly 60 per cent more volume. Moreover, although creating the cavity with an earlier explosion was cheaper than conventional mining techniques, the salt structure

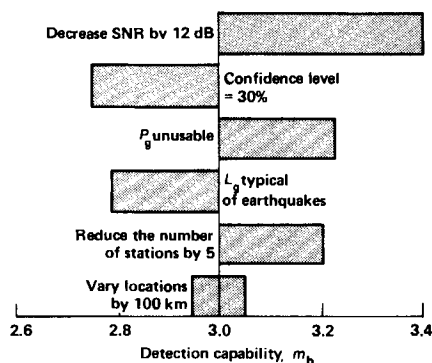


Fig. 3 Sensitivity of the overall detection capability to variations in the assumptions made in calculating the network capability shown in Fig. 4.

around the cavity was thereby weakened (the theory predicts the decoupling factor to be proportional to the rigidity modulus of the surrounding medium).

The significance of these results can be seen when they are combined with the lower limit of seismic detectability. As illustrated in Fig. 1, an m_b measurement of 2.7 now corresponds to a 5 kt explosion (instead of the tamped hard-rock value of 0.1 kt), even if the decoupling factor is only 50. Decoupling factors several times larger are predicted if the cavity can be excavated in a stiffer medium, such as granite⁹.

Although the process has been shown to work, recent objections have arisen on theoretical grounds^{10,11}. Figure 5 illustrates the problem. The spectral variation of the seismic amplitude is plotted, for a fixed yield, for both tamped and cavity-decoupled explosions. In either case it is observed that only modest changes occur as the angular frequency is increased to the neighbourhood of the eigenfrequency, $\omega \approx c/R$, where c is the wave speed and R the elastic radius, after which severe damping occurs. For the cavity-decoupled explosion, this 'corner' frequency is based on the cavity radius, R_D . For the tamped explosion, the radius at which the shock wave attenuates to an elastic wave, R_T , is much larger.

A heuristic explanation for this derives from the fact that the radial momentum increase in a strong explosion is proportional to the square root of the mass engulfed by the blast wave, which is evidently many times larger when the cavity is initially filled with rock than with air. Thus, even though the amplitude of the tamped explosion is 50 or more times greater for $\omega < \omega_T = c/R_T$, it may be much less in the region $\omega_T < \omega < \omega_D = c/R_D$. The decoupling factor can therefore be expected to decrease substantially in this frequency interval. The effect is of no significance for teleseismic monitoring, since frequencies greater than ω_T are effectively attenuated by the Earth. There is some evidence, however, that efficient high-frequency transmission can occur at regional distances in certain geographical regions of importance. Also, Earth noise decreases with increasing fre-

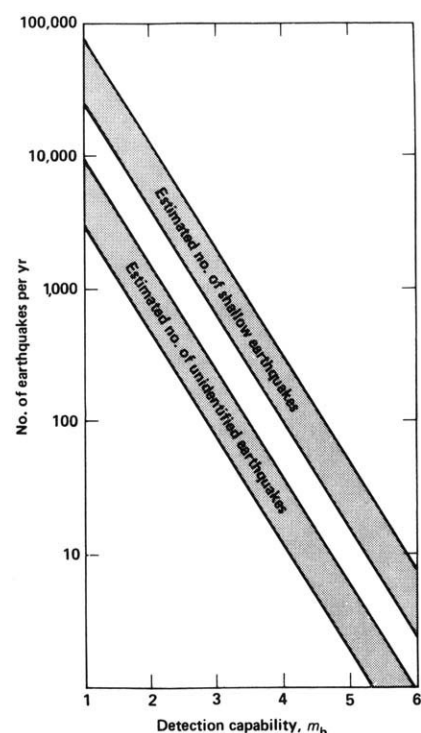


Fig. 4. The average number of shallow seismic events above a given magnitude. 20 per cent of the detected events within 0.5 mag of the overall detection threshold are assumed to be unidentified.

quency, so that increased signal-to-noise ratio (SNR) will aid in signal analysis.

The decrease of the decoupling factor at high frequency has not been confirmed experimentally. The remedy for the problem, if it exists, is to increase the size of the decoupled cavity. However, fully to decouple a 5 kt explosion requires a salt cavity with a radius of 49 m (ref. 12), so that it is not feasible, or at least impractical, to excavate a sphere of the required size. An alternative method of increasing the characteristic size of the cavity is to connect long tunnels to the explosion chamber. These could easily extend beyond distances corresponding to R_T , and could branch and/or be multiply connected. Whether such a system could provide sufficient high-frequency damping is uncertain at present and is an active area of research.

Cavities may also be constructed in a harder rock medium, although this would be more costly and the maximum size may be more limited. Calculations indicate that, for a minimum volume system, the shape of a decoupled cavity in hard rock is much more flexible than in salt, so that tunnel-like excavations could be used¹². These would be smaller in span dimension, but much longer and with less well-characterized seismic radiation patterns. Because hard-rock formations are often located in regions of very active seismicity, cavities excavated there would also seriously complicate the already difficult problem of discriminating between low-yield explosions and earthquakes.

The excavation of large underground cavities would not be hard to disguise from

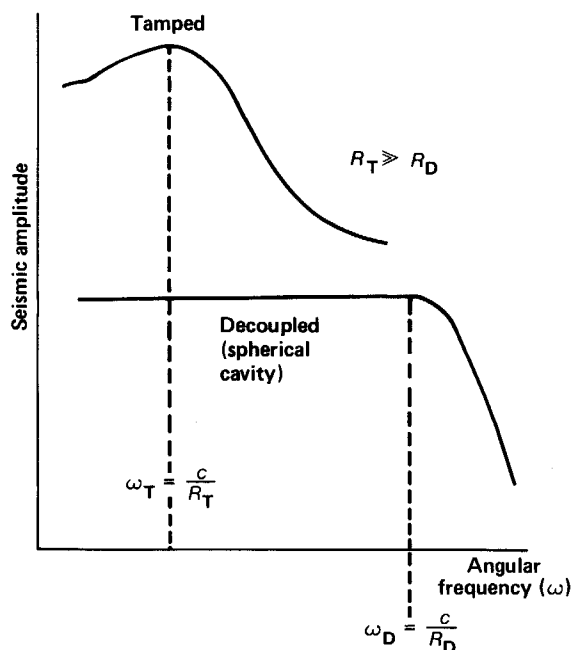


Fig. 5 Spectral variation of seismic amplitude at fixed yield for tamped and cavity-decoupled explosions. The ratio of the tamped to the decoupled amplitude is called the decoupling factor and theoretically decreases at frequencies above ω_T .

satellite observation if the activity were associated with mining or other large-scale industrial undertakings. In salt, the least expensive method is solution mining. A 1970 study indicated that a 50 m radius cavity could be solution-mined in the Tatum salt dome at a cost of less than \$10 million²⁰. By 1977, the cost was put at \$20 million but it was determined that a single cavity might be useful for up to 25 tests, performed over a period of two years²¹. Re-use of a cavity would probably be an important consideration since present-day costs are undoubtedly still higher. With salt, however, there is always the SALMON option of creating a primary cavity with a tamped explosion and then enlarging it in steps by exploding larger and larger devices therein. It is possible that this technique may already have been employed by the Soviet Union in their Peaceful Nuclear Explosion (PNE) programme in which several nuclear explosions have been conducted for the stated purpose of storing the natural gas condensates from Orenburg and Astrakhan fields²².

Conclusions

A total ban on *all* nuclear testing cannot be verified by seismic (or in fact by any other technical) means. What is feasible depends, to some extent, on what is negotiable. For example, if in-country seismic networks are excluded, so that only extraterritorial monitoring can be used, clandestine explosions with yields of up to 10 kt would be possible in appropriate geological formations, even without cavity decoupling. Still larger explosions might go undetected if they were partially decoupled — conducted in cavities too small to allow

full decoupling, but big enough to reduce the seismic amplitude by, say, a factor of 5–10.

The situation changes if internal stations are permitted. In the best case, where a large number of internal seismic arrays are deployed, the probability is high that an evader would be caught attempting explosions with yields exceeding 5–10 kt, even employing fully decoupled cavities. Even with the large internal network, however, smaller explosions would go undetected. A treaty negotiated on the basis of seismic networks alone might even favour the Soviet side, where clandestine testing, in violation of a test-ban treaty, would be much less likely to be reported unofficially than in the United States.

Assuming that a political decision has been made by both sides that it is in their interests to negotiate a new test-ban treaty, how then can any progress be made? An interesting solution has been proposed by R. E. Kidder of our laboratory (unpublished letter). This rationale is that if undetectable nuclear explosions of low yield are possible, then verifiability requires that these explosions should be allowed. Such tests would be sanctioned only at a single designated and seismically well-characterized site in each country, and would be required not to produce a near-field seismic signal in excess of a specified value. The latter condition would require cavity decoupling in order to achieve the maximum yield. (If, instead of a seismic amplitude limit, a yield limit was specified, without accounting for decoupling, there would be no way of seismically verifying that a still larger yield had not been tested in a cavity). The designated sites in each country could be calibrated with radio-chemical probes provided by both sides and also with known charges of chemical explosives.

The advantage of this so-called maximally restrictive verifiable test ban (MRVTB) would be that the present yield limit of 150 kt could be reduced by a factor of 30 with high confidence that neither side was cheating. The incentive for attempting to evade the agreement would be removed for low-yield tests (below 5–10 kt) and the internal seismic network would prevent high-yield clandestine tests, even with full or partial decoupling.

Even with a MRVTB, however, there is no guarantee that isolated clandestine tests could not be made at other locations, to conceal the extent or nature of a test programme. □

1. Statement by the United States Representative (Martin) to the Conference of the Committee on Disarmament; Comprehensive Test Ban, 24 August 1972, *Documents on Disarmament*, 589 (United States Arms Control and Disarmament Agency, (1972).
2. Bolt, B.A. *Nuclear Explosions and Earthquakes*, 67 (Freeman, San Francisco, 1976).
3. Evernden, J.F. *J. geophys. Res.* **75**, 1028-1032 (1970).
4. Hannon, W. J. Jr *Energy and Technology Review*, 50-65 (Lawrence Livermore National Laboratory, 1983).
5. Latter, A.L. & Bethe, H. in *Verbatim Record of the 7th Meeting of Technical Working Group 2, Conference on the Discontinuance of Nuclear Weapons Tests*, Geneva, 2 December 1959 (GEN/DNT/TWG.2/PV.7, 1960).
6. Herbst, R.F., Werth, G.C. & Springer, D.L. *J. geophys. Res.* **66**, 959-978 (1981).
7. Springer, D., Denny, M., Healy, J. & Mickey, W. J. *geophys. Res.* **73**, 5995-6011 (1968).
8. Rawson, D., Taylor, R. & Springer, D. *Naturwissenschaften* **54**, 525-531 (1967).
9. Latter, A.L., LeLevier, R.E., Martinelli, E.A. & McMillan, W.G. *J. geophys. Res.* **66**, 943-946 (1961).
10. Blandford, R.R. *Bull. Seism. Soc. Am.* **72**, S60-S87 (1982).
11. Larson, D.R., *Spherical Wave Propagation in Elastic Media and its Application to Energy Coupling for Tamped and Decoupled Explosions* (Lawrence Livermore National Laboratory Rep. UCRL-52655, 1979).
12. Glenn, L.A. *J. geophys. Res.* (submitted); Lawrence Livermore National Laboratory Preprint No. 90321, (1984).
13. Sykes, L.R. & Evernden, J.F. *Scient. Am.* **247**, 47-55 (1982).
14. Racine, D. & Kiouda, P. *Seismicity of the Salt Areas of Texas, Louisiana, Oklahoma, and Kansas* (Teledyne Geotech Rep. AL-79-3, 1979).
15. Dahlman, O. & Israelson, H. *Monitoring Underground Nuclear Explosions* (Elsevier, New York, 1977).
16. Husebye, C.E. & Mykkelveit, S. *Identification of Seismic Sources — Earthquake or Underground Explosion*, NATO Advanced Study Instit. Ser. 74, Ser. C (Reidel, Boston, 1981).
17. Pomeroy, P., Best, W. & McEvelly, T. *Bull. Seism. Soc. Am.* **72** (6B) S89-S130 (1982).
18. Evernden, J.F. *Bull. Seism. Soc. Am.* **66**, 549-592 (1976).
19. Marshall, P., Springer, D. & Rodean, H. *Geophys. J.* **57**, 609-638 (1979).
20. *Project Payette Final Summary Report* (Fenix & Scisson, Tulsa, Oklahoma, 1970).
21. Grove, L.A., Hubbard, H.W. & Martinelli, E.A. *Feasibility of a Decoupled Test Facility in a Salt Dome* (R&D Associates, Marina del Rey, California, Rep. RDA-TR-104900-013, 1977).
22. Borg, I.Y. *Energy and Technology Review*, 30-37 (Lawrence Livermore National Laboratory, 1983).

Relativity's most elaborate test

A project to test the general theory of relativity is still going after a quarter of a century. The physics is such fun that the travelling may be better than the arrival.

Stanford, California

How do you make into a perfect sphere a lump of fused quartz the size of a ping-pong ball with an accuracy of one part in 10^7 ? How do you suspend that electrically in a cavity hardly any bigger, and set it spinning so as to function as a nearly perfect gyroscope? And then do all this in a satellite at an altitude of more than 500 km? And why, in any case, do you bother?

These questions have for thirty years preoccupied the Stanford-centred group of physicists and engineers designing what must be the most elaborate of all experiments in physics — the measurement of the precession of a nearly-perfect gyroscope moving through the Earth's gravitational field. The declared objective is to test Einstein's theory of gravitation in a way that has not previously been possible. The hidden agenda may be to push a constellation of exacting technologies further than has previously been done.

Since the earliest days of general relativity, it has been clear that a gyroscope should precess (drift) in the gravitational field of a rotating object such as the Earth, but the Stanford project was inspired by the late Leonard Schiff, who calculated explicitly the rate of precession of a gyroscope in free fall around the Earth.

There are two effects, one called geodetic precession and caused merely by motion through the gravitational field, the other due to the angular rotation of the Earth and called motional precession. The first effect is analogous to spin-orbit coupling in the calculations of the energy levels of atoms, the second to the spin-spin coupling between the nucleus and the electrons of an atom which is responsible for the hyperfine structure of atomic spectra.

Both effects are tiny. Geodetic precession should amount to 6.9 seconds of arc a year for a satellite at an altitude of 550 km, motional precession to a mere 0.044 arc-seconds a year. Fortunately, the two effects are at right angles and so may be separated by a single measurement. The objective of the Stanford measurement is to obtain the motional precession to within 1 per cent, which requires an underlying accuracy of 0.3 milliarc-seconds a year — the angle subtended by the width of a human hair at a distance of 10 miles.

The refinements elaborated in the past thirty years have been dictated simply by this goal. The project calls for a gyroscope with an inherent drift-rate nine orders of magnitude less than that of the instruments

now used for the inertial navigation of nuclear submarines. The only hope of achieving such stability is in an orbit about the Earth, for otherwise even the best suspensions will yield an unacceptable amount of drag. The evolution of this project shows what happens when people set outrageously ambitious goals, promising themselves that they will solve whatever unanticipated problems arise. This is how the project has moved from one apparently insuperable obstacle to another:

- The original decision that the gyroscope should be a spinning sphere has not been seriously reconsidered. Fused quartz is used chiefly for strength and homogeneity. Electrostatic suspension is achieved by coating the sphere with a superconducting material (niobium). The space between the sphere and its spherical housing is typically 40 μm . Three pairs of electrodes sputtered on the surface of the housing are used to keep the sphere at the centre of the cavity and also, as capacitors, to sense departures from positions. In the satellite, there will be two pairs of gyroscopes mounted in a line, spinning in pairs parallel and anti-parallel to each other. Making spheres of quartz accurate to one part in 10^7 , which has been done by two special lapping machines at NASA's Marshall Space Flight Center at Huntsville (Alabama), is slightly less difficult than the measurement of departures from roundness, made possible by means of a computerized mechanical stylus built by Rank Taylor Hobson (from Leicester, England) to a design developed at the University of Aberdeen.

- Superconducting skins make essential liquid helium temperatures, which in any case help to reduce random noise. So the spacecraft will be built around a huge Dewar vessel, containing 1,295 litres of liquid helium, like that used successfully in the infrared astronomy satellite IRAS.

- To measure the drift of the gyroscopes relative to the fixed stars, there must be a telescope — one of 400 cm focal length fashioned from fused quartz with folded optics (three coaxial mirror surfaces). The reference point will be the bright star Rigel in Orion.

- Spinning a spherical superconducting sphere cannot be accomplished electromagnetically because of the obdurate diamagnetism of the material, so that the housing is to contain a system of channels for allowing helium gas to flow past the sphere, using friction to get them spinning. The awkward tradeoff that must be made is

between the acceleration of the sphere by the gas jets and its deceleration by gas escaping from the channelled flow. Residual helium is removed by raising the ambient temperature from 1.5 to 3.5 K, whimsically called "baking-out". The intention is that the axes of the gyroscopes should be along the apparent line of sight (uncorrected for aberration) to Rigel, which will entail a complicated sequence of approximations in which the spin axes of the four gyroscopes are alternately adjusted by gas jets and rolling of the spacecraft.

- Sensing the direction in which a nearly perfect spinning sphere is pointing obviously cannot be accomplished optically. Mercifully, F. London showed in 1957 that a superconducting object set spinning should have a characteristic magnetic moment, essentially the external macroscopic magnetic field compensating for the angular momentum imposed by spinning on quantum states of individual electrons. But the London moment is tiny, so it is necessary not merely to exclude all magnetic material from the structure but to isolate the gyroscopic assembly from the external magnetic field. Shifts in the direction of the London moment are to be detected by a pair of superconducting coils and measured with the help of the superconducting devices called SQUIDS.

- Even at 550 km, atmospheric drag will cause the spin axes of the gyroscopes to drift. So it is planned to compensate for drag by placing near the centre of the spacecraft a free-falling metal ball, then using helium jets to keep the spacecraft on the freefall path. This technique was used in the US Navy's navigation satellite TRIAD I to keep the acceleration of that spacecraft below 5×10^{-12} of the acceleration at the surface of the Earth.

The long history of this elaborate project has not been free from trouble, but William Fairbank, professor of physics at Stanford, now thinks the doubters have been won over. NASA has promised space for a proving flight on the shuttle in 1988, and there is a prospect that the satellite may be launched (into an accurately polar orbit) in 1991.

Fairbank says that the outcome may not be just another test of general relativity but perhaps a pointer to the physics that will have to be understood before gravitation can be quantized. If it works, it should help further to make general relativity part of ordinary physics.

John Maddox

Developmental biology

A Rosetta stone for pattern formation in animals?

from Jonathan Slack

FOUR weeks ago in these columns, Gary Struhl described the discovery of a DNA sequence called the homoeo box¹⁻⁵ which promises to be the biological equivalent of the slab of basalt found in 1799 near the Egyptian town of Rosetta by a French officer called Boussard. The Rosetta stone was a decree in honour of Ptolemy V written in Greek and in the ancient demotic and hieroglyphic scripts. It provided the most important clue in the decipherment of the hieroglyphics; eventually scholars were able to read inscriptions on statues, coffins, tombs and temples from all over Egypt. In this article I shall describe some impressive new work on the mechanism of segmentation in the fruit fly *Drosophila*, which was facilitated by the discovery of the homoeo box, and also discuss some of the implications of the presence of the homoeo box in the genome of a wide variety of organisms.

The homoeo box is a protein-coding sequence of 180 nucleotides. There are at least seven copies in the *Drosophila* genome which are sufficiently similar to cross-hybridize under stringent conditions. Six of these seem to be parts of selector genes whose activity codes for different regions of the embryo along the antero-posterior axis; they lie within the Bithorax⁶ or Antennapedia⁷ complexes on the third chromosome. The seventh, also in the Antennapedia region of the chromosome, is part of a gene called *fushi tarazu* (*ftz*) whose mutations belong to another group, disturbing the periodicity of the segmentation pattern while leaving unchanged the non-periodic features of the anteroposterior pattern^{8,9}. Homozygous *ftz* embryos, which die before hatching, have only half the normal number of segmental repeats.

The *ftz* gene has been located on the molecular map of the Antennapedia complex by mapping mutant lesions¹⁰, and now its structure and expression in normal development have been determined¹¹. The gene contains two exons with the homoeo box sequence at the beginning of the second exon. Analysis of embryonic mRNA shows that there is a single transcript of 1.9 kilobases (kb) which is at its most abundant during the first 3 hours of development and declines to zero after about 6 hours. The first visible indication of segmentation in *Drosophila* embryos occurs at about 5 hours (ref. 12), so the *ftz*⁺ gene is active at exactly the right time to have a causal role in the establishment of the periodic pattern.

The really remarkable results come from the study of the spatial distribution of this

mRNA in the early embryo using the technique of hybridization *in situ* of DNA to embryo sections¹³. It should be remembered that in the early stages of insect development there are a number of synchronous nuclear divisions without the formation of cell membranes. At a certain stage most of the nuclei migrate to the periphery of the egg to form a syncytial blastoderm and, a few divisions later, nuclei acquire individual cell membranes to produce the cellular blastoderm — a single cell layer surrounding the inner yolk mass. In *Drosophila*, the syncytial blastoderm is formed after 9 nuclear divisions and the cellular blastoderm after 13. Experiments in which small defects have been induced by laser irradiation¹⁴ show that the embryo itself arises from the region between about 15 per cent and 65 per cent of the length of the egg measured from the posterior pole, the remainder becoming extraembryonic structures. At the cellular blastoderm stage, the region corresponding to each prospective segment is three to four cells wide.

Hybridization *in situ* of *ftz*⁺ DNA first gives a detectable signal after the eleventh nuclear division; it hybridizes to a uniform cylindrical band in the region that will develop into the embryo. Over the next two nuclear divisions, this band breaks down into a pattern of seven stripes, each corresponding to the width of about one prospective segment. The RNA disappears by 4.5 hours, which is about the time that the appearance of neuroblasts gives the first visible indication of segmentation. It is assumed, but cannot be proved at present, that the mRNA in the normal embryos is made in those parts of the repeating pattern missing from the *ftz* mutant.

The implications of this result are far-reaching. First, it provides direct evidence that the ultimate segmental repeating pattern is preceded by a 'prepattern' in which the repeating unit is twice as wide. This is a radically new idea, first suggested by the discovery of 'pair rule' mutants. The prepattern is evidently present in both ectodermal and mesodermal prospective regions of the blastoderm, and so precedes the establishment of germ layers at gastrulation.

Second, it seems that the prepattern is continuous to begin with, but subsequently breaks down into the sevenfold repeat. It is unlikely that this change is the result of the local sorting out of active and inactive nuclei, because of the continuity of genetically-marked clones in the larval hypoderm¹⁵. Both results seem more compatible with a model for segmentation

advanced in 1978 by Kauffman *et al.*^{16,17}, but which, on the whole, has not found favour with other drosophologists. In this model, the embryo is traversed by a sequence of chemical concentration waves which have successively 1, 2, 4, 8 and 16 peaks; this model is claimed to account not only for the formation of the repeating pattern but also for the unique states of determination of the imaginal disc rudiments in each segment.

One of the objections to this model has been that most insects do not form their segments simultaneously but in progression, often from the anterior to the posterior end, with each segment being formed sequentially from a posterior growth zone¹⁸. Nobody wants to believe that the repeating pattern in *Drosophila* is set up by a mechanism completely different from other insects, and yet this is implied by the *in situ* hybridization results with *ftz*⁺ DNA and of the simultaneous appearance of the two-segment repeats.

It is to deal with this sort of pessimistic speculation that the homoeo box returns as a sort of master trump. McGinnis *et al.*³ reasoned that, if the homoeo box is part of several *Drosophila* genes concerned with regional specification, be they selector genes or periodicity genes, then homologous sequences can perhaps be found in the genomes of other animals. And, indeed, they have now been found in *Drosophila virilis*, *Tenebrio* (a beetle), the earthworm, the chick, *Xenopus*, the mouse and the human.

So far, only one of these genes, from *Xenopus*, has been examined in detail¹⁹. A clone called AC1, extracted from a *Xenopus* clone bank using the *Drosophila* homoeo box as an identifying probe, turns out also to contain a protein-coding sequence very similar at the amino acid level to that of *Drosophila*. The appearance of mRNA complementary to AC1 has been studied during early development. There are three RNA species, 1.2, 1.6 and 2.3 kb in size. The 1.2 and 1.6 kb species are first detectable in the late gastrula, a stage at which we expect regional specification to be taking place. The 1.2 kb RNA disappears at the tailbud stage, just as the 2.3 kb species first appears, while the 1.6 kb RNA is present throughout early development.

The implication is that the genes which control regional specification in *Drosophila* have universal functions and that the homoeo box enables them to be identified and extracted from any animal whatever. If so, then we really can anticipate that the problem of pattern formation, one of the deepest mysteries in biology, is on the point of final solution.

What comments can an embryologist make about such results? At first sight, it would be somewhat surprising to find that the mechanisms of pattern formation are really the same in insects and vertebrates. Not only is their descriptive embryology completely different in all respects, but so

is their experimental embryology. Antero-posterior specification in insects occurs in response to a signalling centre at the posterior pole, whereas in vertebrates, the first inductive interaction establishes the three germ layers and anteroposterior specification seems in some way intimately associated with the gastrulation movements²⁰. The homoeo box appears to code for some sort of genetic regulatory element and it is possible that the genes with which it is associated in *Xenopus* have nothing to do with pattern formation. In this context, it is worth noting that McGinnis *et al.*³ report that the homoeo box has not been detected in nematode or sea urchin DNA, and they associate this with the absence of visible segmentation in these creatures. However, only one out of seven copies of the homoeo box in *Drosophila* is associated with a periodicity function (*ftz*), while the other six are associated with selector genes believed to be responsible for antero-posterior specification. Since all animals, whether or not they are segmented, have some anteroposterior differentiation, we might have expected to find the homoeo box in nematodes and sea urchins as well.

Such questions can only be decided by experiment and we await the results of *in situ* hybridization studies with AC1 with bated breath. If AC1 and the other genes containing homoeo boxes have a significant regional expression in the early embryo, we shall really feel there is a deep unity about pattern formation, as proposed by Wolpert²¹ in 1969, which underlies the apparent superficial complexity and diversity of animal development.

For vertebrate embryologists, there will be exciting times ahead. We can expect the inducing factors that appear to control regional specification to function by activating the appropriate combinations of selector genes. If these genes contain the homoeo box, cloned probes will soon be available; then, perhaps, we shall finally be able to find out what is happening during embryonic induction. □

1. Struhl, G. *Nature* **310**, 10 (1984).
2. McGinnis, W. *et al.* *Nature* **308**, 428 (1984).
3. McGinnis, W. *et al.* *Cell* **37**, 403 (1984).
4. Scott, M.P. & Weiner, A.J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4115 (1984).
5. Laughon, A. & Scott, M.P. *Nature* **310**, 25 (1984).
6. Lawrence, P.A. & Morata, G. *Cell* **35**, 595 (1983).
7. Wakimoto, B.T. & Kaufman, T.C. *Dev. Biol.* **81**, 51 (1981).
8. Wakimoto, B.T., Turner, F.R. & Kaufman, T.C. *Dev. Biol.* **102**, 147 (1984).
9. Nusslein Volhard, C. & Wieschaus, E. *Nature* **287**, 795 (1980).
10. Scott, M.P. *et al.* *Cell* **35**, 763 (1983).
11. Kuroiwa, A., Hafen, E. & Gehring, W. *Cell* (in the press).
12. Poulson, D.F. in *The Biology of Drosophila* (ed. Demerec, M.) 168 (Wiley, New York, 1950).
13. Hafen, E., Kuroiwa, A. & Gehring, W. *Cell* (in the press).
14. Lohs Schardin, M. *et al.* *Dev. Biol.* **73**, 239 (1979).
15. Szabad, J. *et al.* *Dev. Biol.* **73**, 256 (1979).
16. Slack, J.M.W. *Nature* **271**, 403 (1978).
17. Kauffman, S.A. *et al.* *Science* **199**, 259 (1978).
18. Sander, K. *Adv. Insect Physiol.* **12**, 125 (1976).
19. Carrasco, A.E. *et al.* *Cell* **37**, 409 (1984).
20. Slack, J.M.W. in *From Egg to Embryo. Determinative Events in Early Development* (Cambridge University Press, 1983).
21. Wolpert, L. *J. theor. Biol.* **25**, 1 (1969).

Jonathan Slack is at the Imperial Cancer Research Fund Mill Hill Laboratories, Burton-hole Lane, London NW7 1AD.

Cosmology

Cosmic strings and galaxies

from Craig Hogan

IN this issue of *Nature*, Kaiser and Stebbins propose an interesting test of the hypothesis that the formation of galaxies and galaxy clusters was induced by the action of 'cosmic strings'. It is only recently that strings have come to the attention of scientists working outside the arcane realms of mathematical physics. In this article, I shall explain what they are.

In modern gauge theories of fundamental interactions, 'the vacuum' is far from being 'nothing'. Rather, it is now recognized as a dynamical object, which can be in different states. The current state of the vacuum affects the properties (such as the masses and interactions) of any particles put into it. Although the vacuum is thought to lie in its ground state — that with the lowest energy — this state has not

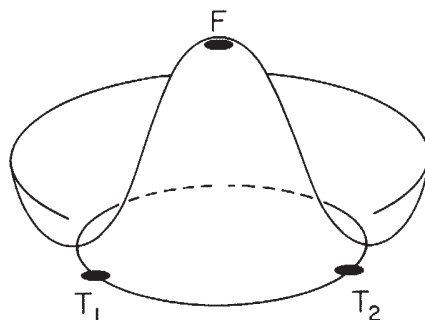


Fig.1 'Mexican hat' potential $V(\phi_1, \phi_2)$ above the (ϕ_1, ϕ_2) plane. At high temperature, thermal effects force the vacuum to the 'false' vacuum at the origin (F), whereas at low temperatures this vacuum is unstable and moves to the 'true' vacuum around the base of the rim. Widely separated parts of the Universe move randomly to different vacua T_1 and T_2 which have the same energy. Any closed path in space which maps onto a loop going around the rim encloses a 'string' where the vacuum is topologically trapped at F . The string itself consists of the region of space where the vacuum differs significantly from its ground state at the rim.

always been the same. Thus, in the early Universe, when the particle component (that is, ordinary matter and radiation) was at a very high temperature, the vacuum adjusted its state, and in doing so modified the properties of the particles, so as to minimize the free energy of the entire system, vacuum plus particles. (That is, the vacuum went into a higher-energy state in order to lower the energy of the hot plasma by an even greater amount.) As the Universe cooled, to keep the entire system at the lowest possible energy at a given temperature, the vacuum had to change, eventually ending up in its present state (which is very nearly the 'true' or zero-temperature vacuum).

It is possible that as the Universe expanded the cooling happened too rapidly

for the vacuum to find its true ground state, and the vacuum was frozen into a ground state with defects. Defects in a three-dimensional space can be zero dimensional (monopoles), two dimensional (domain walls), or one dimensional (strings). The types of defects that may appear depend on the internal degrees of freedom which define the vacuum's state, so they depend on the particular gauge theory under discussion. In general, topologically stable defects form if there is a set of possible ground-state vacua which have the same energy, but which form a curve or surface with nontrivial topology in the internal space describing the vacuum (Kibble, T.W.B. *J. Phys.* **A9**, 1387; 1976 and *Phys. Rep.* **67**, 183; 1983). For example, suppose that the state of the vacuum is described by two parameters, ϕ_1 and ϕ_2 , and that the free energy density of the zero-temperature vacuum is a function $V(\phi_1, \phi_2)$ shaped like a sombrero over the (ϕ_1, ϕ_2) plane (Fig. 1). At high temperature thermal effects force the vacuum everywhere in physical space to the same point at the origin $\phi_1 = \phi_2 = 0$, near the peak of the hat in the vacuum's internal space. As the Universe cools, the vacuum settles down to a point in the circle forming the base of the rim. The lowest energy would be found by lining up the vacuum at all points in space at the same point on the rim, and this is what happens locally. However, widely separated places in the Universe cannot communicate with each other in the time it takes this cooling to occur, and will randomly choose different places around the rim to cool to. The result will be closed paths in space whose corresponding vacua form a loop going right around the rim. If we then deform our path to a point in physical space, the corresponding loop of vacua must somewhere leave the rim of ground states in order also to deform to a point; we have 'trapped' a region of space in the 'false' vacuum at the origin, or created a topological defect. The region trapped in this way is a string in three-dimensional space. Effectively, it is confined to a narrow path in space by the volume energy cost of the false vacuum at $V(\phi_1 = \phi_2 = 0)$ but it cannot shrink away to a point because of the energy cost of having large-amplitude spatial gradients in ϕ .

Such strings are macroscopic objects. In the above example (and in most cases of cosmological interest) they have no ends, and are either infinitely long or in closed loops. They have a longitudinal tension $p = -\mu$, and mass per unit length μ determined by the potential V :

$$G\mu/c^2 \approx (\langle |\phi| \rangle / m_{pl})^2$$

where m_{pl} is the Planck mass and $\langle |\phi| \rangle$ is the radius of the base of the hat rim. The

width of the string (the width of the region where the vacuum differs significantly from its ground state) is so small that the only significant interactions of strings in the present Universe are gravitational.

Strings of this general type have been studied by mathematical physicists for a number of years, but became particularly interesting to astronomers only when unified Yang-Mills theories (also known as grand unified theories) were formulated and enormous values of $\langle |\phi| \rangle \approx 10^{-3} m_{pl}$ were being contemplated corresponding to $\mu \approx 10^{-6} c^2/G$ or roughly a thousand tonnes per fermi. Thus, gravitational effects associated with grand-unified-theory strings would give rise to velocities of order $\sqrt{G\mu} \approx 300 \text{ km s}^{-1}$ comparable with the largest observed astronomical systems. The enormous mass of these strings may be grasped by imagining a closed loop as big as, say, a person; an object orbiting around this loop at a distance of a few metres would have a velocity of order 300 km s^{-1} . Zeldovich (*Mon. Not. R. astr. Soc.* **192**, 663; 1980) and Vilenkin (*Phys. Rev. Lett.* **46**, 1169; 1981) have studied the evolution of the network of strings formed in the early Universe, and suggest that the gravitational effects of cosmic strings are actually responsible for creating galaxies and clusters.

The effect discussed by Kaiser and Stebbins on page 391 is a good illustration of the peculiar nature of the interactions of vacuum strings. The mass of strings is not like the mass of ordinary matter, because it is due entirely to the string tension. Thus, if we grab a loop and stretch it out, its mass increases, proportional to its length; this new energy will have come from the work we did against the string's tension. The gravitational interactions of strings are not the same as those in an ordinary massive rod. Spacetime containing only a straight string is 'conical', so it induces no tidal forces or gravitational attraction. If an infinite stationary straight string were passing through the room, it would not affect anything — you would feel no gravitational pull of any kind (this would not be the case if the string were curved). If, on the other hand, you walked through the string with velocity V , you would find that the parts of your body which had passed on opposite sides of the string would be moving towards each other with velocity $8\pi V \times (G\mu/c^2)$; the string leaves a 'wake'.

Because of the enormous tension, $p = -\mu$, of cosmic strings, disturbances travel along them at the speed of light, and therefore strings which are not perfectly straight generally tend to move transversely at near the speed of light. If one of these fast strings were to pass through the room, you would certainly notice it; the parts of the room on opposite sides of the string's swath would suddenly be moving towards each other with velocity $\sim 8\pi c(G\mu/c^2) \approx 4 \text{ km s}^{-1}$. Unless it actually passed through your body, however, you would still feel nothing (except perhaps some anxiety)

until something from the other side actually hit you in the face.

A similar wake forms behind strings in the cosmic background radiation, and this is the effect discussed by Kaiser and Stebbins. The observer's telescope and the last scattering surface find themselves moving towards each other in the wake on the trailing side of a transversely-moving string, so the temperature is seen to change abruptly by a Doppler factor $V/c \sim 8\pi G\mu/c^2$ between the leading and trailing sides of the string. The formation of this wake is dissipative and produces a weak viscous drag on the string.

Clearly it is of great interest to know whether these exotic relics are indeed largely responsible for the largest aggregations of matter in the Universe, both from the standpoint of astronomers, who have been struggling for years with the origin of large-scale structure, and from that of physicists, who would be delighted to find observable phenomena related to fundamental physics concerning the internal structure of the vacuum at very

short distances, inaccessible to accelerators but pertaining to the early Universe. The distinctive nature of the microwave anisotropy predicted by Kaiser and Stebbins is probably the cleanest and most direct way of searching for the effects of strings. Other more indirect effects may be discerned, such as gravitational lensing of quasars by string loops, or a cosmological gravitational wave background produced by rapid oscillations of string loops, which might be detectable using timing measurements of the millisecond pulsar 1937+215. One of the most attractive features of strings is the fact that they have so many distinctive and unique observational side effects; cosmologists might find strings to be unusual among the proposed relics of the early Universe in that they can be ruled out observationally before they are superseded by theoretical fashion. □

Craig Hogan is in the Department of Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125.

Molecular botany

Plant cells under stress

from C.A. Cullis

PLANTS are subject to a wide variety of environmental and biological stresses throughout their life cycle and some species have the ability to adapt to often harsh and variable environments. Different conditions of stress may induce common responses, such as the enhancement of ethylene production, and a large number of stress responses have been characterized. The way in which these specific changes are controlled and the molecular messages involved in that control are generally unclear, but contributions to a recent meeting* showed that clarity is forthcoming.

The response of soybean to heat shock has been studied extensively (J.L. Key *et al.*, University of Georgia). In response to temperature increases from typically 28°C to 40°C a class of heat-shock genes is activated in this plant. The accumulation of the corresponding proteins is not necessarily correlated with a reduction in the synthesis of other proteins (E. Vierling and J. Key, University of Georgia) although towards the highest temperatures tolerated (45°C) they are the major class of proteins synthesized. The presence of the heat-shock proteins is strongly correlated with the acquisition of thermal tolerance and the thermal protection afforded by the proteins is associated with their localization within the cell (C.Y. Lin *et al.*, National Taiwan University; P. Cooper *et al.*, University of Illinois). Pre-adaptation of maize seedlings to heat shock allows them

to recover fully from the severe damage to the photosystem II complex of chloroplasts caused by heat shock (C. Arntzen *et al.*, Michigan State University — D.O.E. Plant Research Laboratory). This recovery is directly correlated with the rates of synthesis of new thylakoid polypeptides.

A number of stress-induced substances, including the enzymes of the phenylpropanoid pathway, hydrolases, protease inhibitors and phytoalexins, have been characterized. In cultured parsley cells, the enzymes of the phenylpropanoid pathway can be induced by various stresses (K. Hahlbrock *et al.*, Max Planck Institute, Cologne). Induction of several of these enzymes is controlled at the level of gene transcription but not indiscriminately so. Thus UV light increases the transcription of the phenylalanine ammonia-lyase gene more than that of the 4-coumarate: CoA ligase gene, though with the same time course, whereas the opposite is found when the inducing agent is 'elicitor' — polysaccharide fragments of cell walls which are responsible for the responses of plants to pathogen attack or wounding. Heat shock can override the stimulation of phenylalanine ammonia-lyase synthesis by UV light or elicitor, irrespective of which stress is applied first, and the elicitor overrides the UV light effect. Thus, there appears to be a hierarchy of stress response.

Chitinase and β -1,3-glucanase activities can be specifically induced either by fungal infection (and elicitor) or by ethylene (T. Boller, University of Basel, Switzerland). However, the suppression of ethylene pro-

* ARCO Plant Cell Research — UCLA Symposium on 'Cellular and molecular biology of plant stress', organized by J.L. Key and T. Kosuge, Keystone, Colorado, 15–20 April 1984.

duction in response to stress does not alter the level of production of either enzyme in immature pea pods by the pathogen (and elicitor), indicating that ethylene and the pathogen are independent stimuli of these two hydrolases.

Production of proteinase inhibitors I and II of tomato is greatly stimulated by wounding, although there is a low level of message in the unwounded plant (C.A. Ryan *et al.*, Washington State University, Pullman). Both the cDNA and genomic clones for these genes have been obtained from tomato and potato. Again polysaccharide fragments of cell walls are active as inducers. They are also potent elicitors of the castor bean and pea phytoalexins (C.A. West *et al.*, University of California, Los Angeles). The accumulation of the soybean phytoalexin, glyceollin, is preceded by increases in the activity of a number of enzymes including those involved in its synthesis (H. Grisebach, University of Freiburg). Infection of soybean with *Phytophthora megasperma* f.sp. *glycinea*, or treatment with a glucan elicitor from this fungus, leads to *de novo* enzyme synthesis which is controlled at the level of gene transcription. Using a radioimmunoassay for glyceollin it was shown that in soybean roots infected with a compatible race of the fungus, glyceollin accumulated only in the epidermal layer, whereas with an incompatible race glyceollin also accumulated to a high level in the root tissue, thus supporting the role of this phytoalexin in the plant's defense reaction.

Cell-wall polysaccharides clearly play a part in the regulation of gene expression in response to pathogen attack. But they may also have diverse regulatory functions in growth and development and may be an important class of regulatory molecules (P. Albersheim *et al.*, University of Colorado, Boulder). When most of molecular biology is devoted to proteins and nucleic acids it is timely to consider that carbohydrates may be more than just structural components of the cell. □

C.A. Cullis is at the John Innes Institute, Colney Lane, Norwich NR4 7UH.

100 years ago

THE employment of acupuncture by Chinese doctors forms the subject of an article in one of the last numbers of the *North China Herald*. A native public writer not long since claimed that a skilful physician in this department of medicine could cure such diseases as imbecility, fits, cholera, &c. The principle of cauterisation is simply that of counter-irritation; and the English writer bears personal testimony to its efficiency in the case of a slight sunstroke, although the operator was a simple Manchu peasant, and instrument a couple of copper coins. Very extraordinary cures are attributed to acupuncture by the Chinese. It is first performed in the hollow of the elbow of each arm. If the puncture draws blood there is no danger, but if no blood appears the case is regarded as very grave. But before abandoning the sufferer, puncture of the abdomen is tried.

From *Nature* 30, 324, 31 July 1884.

Bacterial motion

Surprises from halobacteria

from Michael Spencer

INTEREST in the bacteria that thrive in saturated salt solutions has progressed far beyond the original desire to arrest the spoilage of salted meat. It is now clear that halobacteria show some unique features that give them an unusual fascination for molecular biologists. The best known of these is the purple membrane that functions as a light-driven proton-pumping system for driving the synthesis of ATP. Solar-energy buffs have even suggested that the system might usefully be harnessed to the generation of electricity (Singh, K. & Caplan, S.R. *Trends biochem. Sci.* 5, 62; 1980). As part of their evolutionary adaptation to life in hot shallow salty pools, halobacteria have also developed a double linkage between light-sensitive membrane proteins and the flagella: one system drives them towards regions of high-intensity visible light, while the other helps them to avoid dangerously high levels of ultraviolet radiation (Hildebrand, E. & Schimz, A. *Photochem. Photobiol.* 38, 593; 1983). It is in the actual operation of the flagellar motors that some surprising differences from other bacteria have now emerged.

The saga of flagellar motion (do they wiggle or actually go round and round?) has been running for many years, and progress with more common species was summarized only recently in these columns (*Nature, News & Views* 309, 404; 1984). The evidence strongly favours the idea that each flagellum has an independent and reversible rotary motor, powered by a proton gradient. In many bacteria the flagella operate in bundles which rotate synchronously, but there is a topological restriction on the sense of rotation: if the flagella are coiled into left-handed helices the bundle can only rotate in a counter-clockwise direction when looking towards the body, while right-handed helices can only rotate clockwise. Halobacteria (which also have bundled flagella) appear to break this rule, and in a new paper Alam and Oesterhelt (*J. molec. Biol.* 176, 459; 1984) discuss how this might be explained.

The reason for the normal restriction has been analysed in detail by Macnab (*Proc. natn. Acad. Sci. U.S.A.* 74, 221; 1977), but can easily be demonstrated by twirling two wire helices between the finger and thumb. For one direction of rotation, two or more helices can rotate indefinitely with apparent propagation of helical waves away from the body, whereas with reversed rotation a state of 'jamming' rapidly develops which stalls the engine. In many species the resulting stress leads to a transient reversal of helix sense which makes the bundle fly apart, followed by a period of 'tumbling' and then a reversion

to normal rotation. The result is that the bacterium sets off in a new randomly determined direction after each reversal. Reversals occur spontaneously even in the absence of stimuli, and control by receptors of external stimuli is exercised by varying the interval between reversals: when moving in a favourable direction, reversals are much rarer.

Halobacterium halobium, the organism studied by Alam and Oesterhelt, does not operate like this. In contrast to the frenetic activity of *Escherichia coli*, which darts about at a rate of $20 \mu\text{m s}^{-1}$, *H. halobium* moves towards its goal at the more stately pace of $2 \mu\text{m s}^{-1}$; and when subjected to an unfavourable stimulus it simply pauses briefly before moving backwards along the same line. Alam and Oesterhelt found to their surprise that there was no reversal in the sense of the helix, which remained right-handed throughout; all that happened was that the flagella rotated in the opposite direction, and yet the bundles never flew apart. The authors therefore asked themselves "whether each filament has an individual motor, whether all the filaments are mounted on a single motor plate, or whether halobacterial mobility is dependent at all on a rotatory motor".

The last hypothesis, which would nowadays be regarded as heretical, was discarded after experiments in which bacteria were observed to go round and round when their flagella were tethered to a glass surface with antibodies. Over the remaining two possibilities the authors rather hedge their bets. On the one hand, electron microscopy shows the members of a bundle apparently penetrating the cell surface in a tight coil, which would seem to exclude the insertion of each flagellum into a separate motor of the kind found in other bacteria. On the other hand, the cultures contain large numbers of detached flagella which aggregate spontaneously into long 'super-flagella', and it is possible that a motile bundle is simply one actively-driven flagellum with a number of inactive partners stuck to it. Alam and Oesterhelt suggest that further study of the system at lower salt concentrations, at which the morphology of the cell body remains unchanged but super-flagella are known to disaggregate, might provide the answer. They also draw attention to another interesting difference from other bacteria, in that *H. halobium* flagella show a much more complicated pattern of proteins on gel electrophoresis. These freaks of nature may have more puzzles in store for us yet.

Michael Spencer is in King's College Department of Biophysics, 26-29 Drury Lane, London WC2B 5RL.

Cell biology

More red than dead

from Walter Gratzer

THE opinion has been too often voiced that the mature mammalian red cell, having shed its nucleus, ribosomes and cytoplasmic cytoskeleton, is no more than a floating corpse, fit for the attentions of nothing better than a pathologist. Yet more is now known about its membrane and the proteins that control its properties than is the case for any other cell, and the fact remains that if the mechanisms by which such features as cell shape and lipid asymmetry are regulated cannot be discovered in the red cell, they are likely to remain perpetually obscure.

Among the facts that practitioners in the field would most ardently like to explain are the following: first, the characteristic shape of the cell, the biconcave disc, is under metabolic control. When the cytoplasmic ATP pool is used up the discocyte develops protruberances on its surface and turns into an echinocyte. This is then transformed into a spiculated sphere and the projections are later lost as microvesicles, leaving a smooth spherocyte behind. Up to a quite advanced state of echinocytosis the changes are reversible, if metabolites are added to the medium and the cell is allowed to resynthesize ATP. These effects can be reproduced in ghosts (or empty resealed membranes) and they require ATP uniquely: non-hydrolysable analogues will not serve. Secondly, the four major phospholipids of the membrane are asymmetrically distributed, and remain so for the lifetime of the cell; in particular the two aminophospholipids, phosphatidylserine (PS) and phosphatidylethanolamine (PE), are essentially confined to the inner leaflet of the bilayer. It is, as will be revealed, a new and remarkable development that ATP seems also to be required to maintain this segregation and that the two phenomena — asymmetric lipid distribution and metabolic shape control — may be linked by a common mechanism.

What has long been clear is that the function of ATP in conserving the discoid form is not related to the exclusion of calcium ions, nor to the preservation of ionic concentration gradients of any kind. Until recently it was widely held that the key to its action was phosphorylation, some said of proteins — especially spectrin, the major structural constituent of the membrane-associated cytoskeleton — and others lipids. It later emerged, however, that the transformation to echinocytes on depletion of ATP was largely complete before much of the spectrin had been dephosphorylated, and although the argument is far from over, the protein phosphorylation theory has undoubtedly been badly savaged by this and a range of other arguments (Patel, V.P. & Fairbanks, G.

J. Cell Biol. **88**, 430; 1981).

So what alternatives remain? One is based on the turnover of lipid phosphorus, originally envisaged by Michell and Allan (and now much discussed in relation to malignant transformation). To understand how this might operate, it should be recalled that echinocytosis is conceived by the bilayer-couple model (which is alive and well, despite some recent assaults) to result from an expansion in area of the outer leaflet of the lipid bilayer by no more than a fraction of a per cent — the mathematics has been done — whereas a similar expansion of the inner leaflet leads to a kind of eversion of the cell, with the formation of a cup-shaped stomatocyte. The metabolic model has now been given its most coherent expression by Ferrell and Huestis (*J. Cell Biol.* **98**, 1991; 1984). It is based on the behaviour of minor membrane components, the mono- and diphosphoinositides and phosphatidic acid. When their fuel runs out, ATP-dependent kinases can no longer keep these components phosphorylated in face of continuing phosphatase activity, and the phosphoinositides, which are probably located entirely in the inner leaflet, are hydrolysed to phosphatidylinositol. Dephosphorylation diminishes the area occupied by the molecules and may also allow them to redistribute between the leaflets. On both counts the area of the inner leaflet is reduced relative to the outer. At the same time phosphatidic acid is converted to diacylglycerol, with similar consequences. The estimated ensuing area imbalance of 1 per cent or so could induce echinocytosis, which indeed closely accompanies the dephosphorylation process. The evidence is circumstantial, but quite compelling.

The distribution of the major phospholipids remains unchanged during echinocytosis, and it has been surmised that it is stabilized by interaction of the bilayer with the membrane cytoskeleton. *In vitro* at least, spectrin will bind PS; moreover, when the spectrin is oxidatively cross-linked *in situ*, as Haest and Deuticke have shown, PS and PE rapidly appear in the outer leaflet. Other kinds of affront to the cell evidently have a similar effect, as in malaria, or in sickle-cell anaemia, when reversible sickling occurs. Most strikingly, when the sickling is reversed by deoxygenation, the lipids forthwith return to their normal disposition.

It has now been found that under the right circumstances extraneous phospholipids can be taken up by intact cells and shuffled between the leaflets in a matter of hours, at least in the case of PS and PE. This has been shown in an admirably wrought study by Seigneuret and Devaux

(*Proc. natn. Acad. Sci. U.S.A.* **81**, 3751; 1984), who attached spin labels to phospholipids at the end of one hydrocarbon tail. The accessibility of the label at the outer surface of a cell that has absorbed the lipid can then be assessed from the extent to which the spin signal is annihilated after addition of a reducing agent to the cell suspension. PS and PE, it transpires, largely disappear into the inner leaflet within a few hours, whereas phosphatidylcholine equilibrates much more slowly. The most remarkable discovery, however, is that the passage from the outer to the inner leaflet depends on the presence of Mg-ATP (and again a non-hydrolysable analogue will not do) inside the cell. The immediate result of incorporation of excess lipid into the membrane is crenation of the cell, as the outer leaflet expands. The echinocytes then spontaneously give place to discocytes and/or stomatocytes as the lipids redistribute and the imbalance between the leaflet areas is relieved or partly inverted. When there is no ATP in the cell, the lipids do not redistribute and echinocytosis persists.

Seigneuret and Devaux infer from all this that the lipid asymmetry of the normal cell results from a continuous ATP-dependent transport of PS and PE from the outer to the inner leaflet, opposed by a slow passive back-diffusion against the concentration gradient. This active transport, they suggest, maintains the area balance between the two leaflets that the geometry of the discocyte demands. For this pleasingly simple model of shape control to be upheld, its relation to an assortment of observations, all indicative of an intrinsic preference of PS and PE for the inner leaflet, will have to be carefully examined.

Recently, for example, Raval and Allan (*Biochim. biophys. Acta* **772**, 192; 1984) have found that lipid asymmetry, at least as to PS, persists indefinitely in spectrin-containing microvesicles, shed from calcium-loaded red cells. Perhaps the most convincing evidence that PS and possibly PE do indeed spontaneously seek the inner leaflet of the red cell membrane, and are not merely trapped there, is to be found in a further study by Haest and Deuticke and their co-workers (*Biochim. biophys. Acta* **882**, 328; 1984). They show that lysophosphatidylserine (which lacks one of the two hydrophobic tails present in the parent lipid and can therefore move freely between the leaflets) homes in on the inner leaflet (with a preference of 4:1 over the outer), whereas lysophosphatidylcholine has a slight preference for the outer leaflet.

And what, for that matter, are we to make of the observation (Jinbu, Y. *et al.* *Biochim. biophys. Acta* **773**, 237; 1984), from no less a laboratory than that in which metabolic shape control of red cells was discovered a quarter of a century ago? When the cytoskeleton is cut adrift from the membrane by proteolysis of the primary retaining protein, ankyrin (which can be done with no visible damage to any-

thing else), the shape becomes immutable, at least until the membrane breaks up, unperturbed by metabolic depletion or a crenating drug. Why, what else but that the shape is after all controlled by the cytoskeletal complex. And here too some reappraisal has been called for with the discovery in the red cell of more of the protein components characteristic of contractile systems (Fowler, V.M. & Bennet, V.J. *biol. Chem.* **259**, 5978; 1984).

So we have moved within the year from poverty to riches, to at least three entirely self-consistent explanations of how metabolic shape control operates. No matter that they are almost certainly mutually incompatible, perhaps even all wrong: they will undoubtedly inject good cheer and even perhaps a cautious optimism into toilers in the field. □

Walter Gratzer is in the MRC Cell Biophysics Unit, 26-29 Drury Lane, London WC2B 5RL.

Sequencing techniques

A close look at the genome

from P.F.R. Little

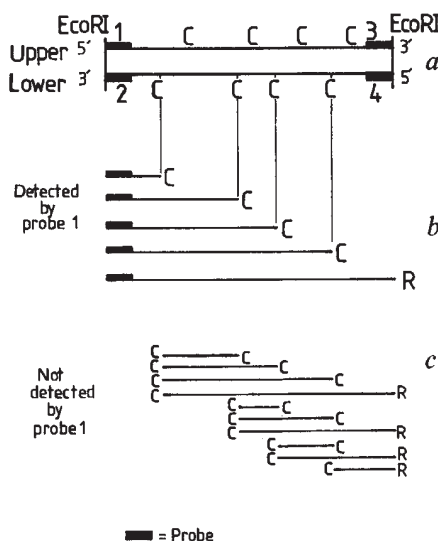
A recent paper¹ in *Proceedings of the National Academy of Sciences U.S.A.* from George Church and Walter Gilbert of Harvard University and Biogen describes a remarkably powerful new method, named genomic sequencing, which provides us with the most accurate view yet of a DNA sequence in a chromosome. It can be used to study methylation patterns directly in uncloned DNA and to analyse DNA sequences associated with protein in chromatin.

The method combines the chemical DNA-sequencing procedures of Maxam and Gilbert² with detection of DNA sequences by Southern blotting³. Figure 1 outlines the procedure by which a DNA molecule can be directly sequenced from unfractionated mouse DNA. It illustrates the sequencing of cytosine (C) residues but similar rationale and procedures are used for sequencing adenine (A), guanine (G) and thymine (T) bases.

The first step is complete cleavage of total cell DNA by a restriction enzyme; Fig. 1*a* shows a double-stranded molecule generated by digestion with *EcoRI*. This is followed by partial cleavage at C residues by hydrazine and fractionation of single-strand DNA fragments on 6 per cent polyacrylamide gels. The DNA is then transferred electrophoretically to nylon membranes and fixed by UV irradiation, and specific DNA sequences are detected by hybridization to a ³²P-labelled nucleic acid probe, as in conventional Southern blots. If we consider only the lower strand of the DNA molecule in Fig. 1*a*, then partial C cleavage will generate the single-strand fragments in *b* and *c*. The fragments in Fig. 1*b* can be detected by hybridization with probe 1 and will show up on an autoradiograph as a set of DNA fragments whose sizes correspond to the distance between the left-hand *EcoRI* site and the position of each C residue. The fragments in Fig. 1*c* are not homologous to probe 1 and so are not detected. The concept is no different from using end-labelled DNA in more conventional sequencing experiments. The novel feature, of course, is that the hybridization step enables one set of fragments to be picked out from the 10⁶

other *EcoRI* fragments in the mouse genome.

In the figure, the DNA detected by probe 1 contains no C residues — a rather unlikely event. What is the effect of cleavage at C residues within the DNA detected by the probe? The result would be fragments similar to those in Fig. 1*b*, but terminating at both ends in C residues, and these would generate on the autoradiogram a second set of fragments, starting at the internal cleavage site. This is not a problem if the frequency (*P*) of C cleavage is small. Since they all terminate on the left with an *EcoRI* site, the fragments in Fig. 1*b* will be generated at a frequency of *P*; by contrast, fragments terminating at a C at both ends will be generated at a frequency of *P*². Thus, internal C cleavage will generate only small amounts of fragments and, consequently, a faint band pattern. Church and Gilbert are able to show that cleavage



DNA fragments detected by genomic sequencing. Probe 1 is the DNA from the upper strand and is homologous to the lower strand. *a*, Double-stranded DNA molecule resulting from digestion of total cell DNA with *EcoRI*. Partial cleavage at C residues by digestion with hydrazine generates two sets of fragments, one of which will hybridize with probe 1 (*b*) and another which will not hybridize (*c*).

at an average of 1 per 500 nucleotides is optimal. The nature of probe 1 is also important. If it is too big, with too many internal Cs, then the cumulative effect of the alternative sequence sets will tend to obscure the primary set. Similarly, if the probe is too short it will hybridize non-specifically and generate other fragment sets. Church and Gilbert indicate that probes of 100–200 nucleotides 'work well'.

A, G and T bases are sequenced in a similar way and the complete sequence can be read directly from the hybridization autoradiograph. An advantage of this method is that filters can be rehybridized, for example with probe 2, 3 or 4. Probes 2 and 4 can be used to sequence the upper strand and probes 1 and 3 the lower. The procedure is remarkably sensitive: as little as 3 fg can be detected using highly radioactive RNA probes made by SP-6 RNA polymerase (reviewed in ref.4) or single-strand DNA probes.

Church and Gilbert have used their new method to study differential methylation of C residues round the mouse immunoglobulin *C_μ* gene in several expressing and non-expressing cell lines. This is possible because 5-methylcytosine is relatively insensitive to hydrazine cleavage, so the appropriate fragments are absent or under-represented in genomic sequences. The method allows the analysis of all 5-methylcytosine bases, not just those that fall within the recognition sites of particular restriction enzymes, and should result in a much clearer understanding of the inverse correlation between methylation and gene activity.

More importantly, perhaps, genomic sequencing can be used to detect specific DNA sequences bound by protein. The protein protects the DNA against chemical or enzymatic attack ('footprinting') and a characteristic 'hole' in the sequence results (see ref.5 for recent advances). This method was originally developed using cloned DNA sequences to study lac and λ repressor-operator and RNA polymerase-promoter interactions. Genomic sequencing can be used in a similar fashion and Church, in collaboration with Anne Ephrussi in Tonegawa's laboratory at MIT, has recently detected protein(s) interacting with the immunoglobulin enhancer at two distinct sites (personal communication).

For the future we can anticipate the use of non-radioactive probes and Church is exploring methods for rapid DNA sequencing. The results of more general application of genomic sequencing are awaited with much interest. □

1. Church, G. & Gilbert, W. *Proc. natn. Acad. Sci.* **81**, 1991 (1984).

2. Maxam, A.M. & Gilbert, W. *Proc. natn. Acad. Sci.* **74**, 560 (1977).

3. Southern, E.M. *J. molec. Biol.* **98**, 503 (1975).

4. Little, P.F.R. *Nature* **309**, 191 (1984).

5. Becker, M.M. & Wang, J.C. *Nature* **309**, 682 (1984).

P.F.R. Little is at the Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB.

Cretaceous-Tertiary extinctions

Physical evidence of impact

from Richard A. F. Grieve

SINCE the original reports four years ago of abnormal abundances of iridium and other siderophiles at the Cretaceous-Tertiary (K-T) boundary, there has been intense interest in the possibility that they represent the geochemical signature of a large impact event that is causally related to contemporaneous mass faunal extinctions. Evidence of equivalent anomalies at boundary sites throughout the world shows they are a global phenomenon. In the meantime, there have been detailed searches for physical evidence of large-scale impact at the K-T boundary. Such evidence is now forthcoming, which makes it virtually certain that such an impact did take place.

Some of the physical evidence is reported by Jan Smit and Frank Kyte on page 403 of this issue of *Nature*. They describe sand-sized microtektite-like spherules at two boundary sites in Italy. There are three types: sanidine-, glauconite- and magnetite-bearing. Although the chemistry and texture of much of the spherule mass have changed over geological time, the magnetite grains in the spherules are believed to have retained their original characteristics. They are enriched in iridium and have approximately chondritic abundances of other siderophiles. In texture, they resemble rapidly-crystallized high-temperature liquids. Smit and Kyte conclude that they are the result either of shock-melting of mafic, possibly oceanic, crustal material or of atmospheric melting and ablation of a mafic projectile.

Spherules from the K-T boundary have been reported previously at Caravaca, Spain (*Nature* 292, 47; 1981), at another Italian site and in DSDP hole 465A in the central Pacific (*Geology* 11, 668; 1983). Similar occurrences in DSDP hole 577A from the North Pacific, reported by Kyte and Smit, indicate that, like the geochemical anomaly, these spherules are not a local phenomenon. The suggestion that the magnetite-bearing spherules originate as melted oceanic crustal material is supported by isotopic data (*Earth planet. Sci. Lett.* 60, 155; 1982, and 64, 356; 1983) that identify a component of oceanic crustal material, considered to be impact fallout, in the boundary layer.

Although the altered character of the boundary-layer spherules puts the interpretation of their impact origin in some doubt, there is independent physical evidence of impact at the K-T boundary. Bohor *et al.* (*Science* 224, 867; 1984) recently recovered quartz grains with microscopic planar deformation features and lowered refractive indices from the boundary layer in Montana, USA. These

features are part of a suite of features known collectively as shock metamorphic features, which are unequivocal evidence of impact. The orientations of the planar features are indicative of shock pressures of scale 15 GPa (150 kbar). Shocked quartz has now been reported from sites in New Mexico, in Spain and from two sites in both Denmark and Italy. So it seems that the occurrence of shocked quartz, too, is not a local phenomenon.

Details of the K-T event are still open to interpretation, however, and the evidence itself raises some questions. The spherules and the isotopic data suggest an oceanic impact. On the other hand, the shocked quartz could be interpreted in terms of a continental impact. In the Montana section, shocked quartz constitutes only 0.01 per cent of the boundary claystone. On the basis of isotopic data, there is an estimated 15–20 per cent terrestrial fallout material in the marine section at Caravaca, Spain. Assuming 0.01 per cent quartz for the boundary layer, this gives a quartz component of approximately 5 per cent in the terrestrial detritus. Perhaps the data can be reconciled by postulating an impact involving oceanic crust and a small component of overlying continentally-derived sedimentary materials. Alternatively, one could invoke other more *ad hoc* explanations involving a number of impacts and thus targets.

The question of the target involved in the K-T event has a bearing on the mechanism of the impact-induced extinctions. An oceanic impact might induce short-term elevated temperatures through a greenhouse effect from increased water vapour in the atmosphere. On the other hand, a continental impact would conform more closely to the original hypothesis of Alvarez *et al.* (*Science* 208, 1095; 1980), which suggests reduced photosynthesis as a result of obscuration by impact ejecta dust.

With the strong evidence of impact at the K-T boundary, there is an understandable desire to relate other mass extinction events to large-scale impact. The Frasnian-Famennian extinction event, for example, has been suggested as a likely candidate. A number of recent papers have linked the extinction record with periodic extraterrestrial phenomena, including cometary showers (*Nature* 308, 709; 1984). If correct, this has far-reaching conclusions. These postulated links should be viewed with caution, however, as they are based on models and statistical analyses of small samples. They offer no geochemical or physical evidence of impact affecting the Earth on a global scale. If the K-T event has abundant physical evidence but still lacks a 'smoking gun' in the form of an impact structure of commensurate age and size, it is not clear whether even 'shots have been fired' at these other extinction events. Careful and detailed work must continue in searching for and investigating the evidence for large-scale impact in the biostratigraphical record. □

Richard A.F. Grieve is in the Earth Physics Branch, Department of Energy, Mines & Resources, Ottawa, Canada K1A 0Y3.

Astronomy

IRAS circular 13

The source name consists of four parts: (1) the letters 'IRAS' to indicate the origin; (2) the right ascension (RA) in hours and minutes, seconds omitted; (3) declination (Dec) in decimal degrees, multiplied by 10 and then truncated (i.e. +32 deg 42.3 arc min => +327); (4) an appendix starting with 'P' and followed by the number of the circular. This appendix stresses that the data are preliminary. Position is given at Equinox 1950.0. Measurements were made between epochs 1983.1 and 1983.8. am, arc min. *Source is extended at this wavelength; a point source flux is not quoted.

Source IRAS	RA h min s	Dec deg am	Flux densities (Janskys)			
			12 μ m	25 μ m	60 μ m	100 μ m
0532+098P10	05 32 23	+09 53.5	1.4	7.1	2.2	<4
0536-026P10	05 36 14	-02 37.6	5.0	15.1	<10	<15
0919-453P13	09 19 28	-45 18.1	47	38	6.0	<4
0944-478P13	09 44 51	-47 48.0	38	46	11	<4
0945-472P13	09 45 02	-47 16.7	41	37	9.5	<3
0947-462P13	09 47 06	-46 17.5	3.2	4.4	1.5	<2
0957-313P13	09 57 52	-31 18.7	0.42	0.91	9.0	21
1012-286P13	10 12 24	-28 37.4	<1.3*	<4*	14	39
1013-413P13	10 13 53	-41 18.4	<0.2	<0.7*	4.2	8.4
1021-284P13	10 21 57	-28 28.5	<0.2	1.0	5.4	7.1
1029-396P13	10 29 24	-39 42.0	<0.8*	<1.8*	7.7	18
1248+482P13	12 48 22	+48 12.3	<0.2	0.57	5.7	9.8
1252+468P13	12 52 20	+46 48.1	<1*	<1.4*	5.5	18
1303+419P13	13 03 34	+41 59.4	<0.5	<0.3	2.0	5.5
1309+469P13	13 09 03	+46 58.0	<0.2	<0.3	3.2	7.2
1323+435P13	13 23 04	+43 31.5	0.41	0.68	7.2	16
1325+479P13	13 25 25	+47 54.7	<0.2	<0.2	2.0	3.3
2002+320P10	20 02 38	+32 04.5	2.5	16	52	86

Influence of degassing on oxidation states of basaltic magmas

E. A. Mathez

Department of Geological Sciences, University of Washington, Seattle, Washington 98195, USA

Intrinsic oxygen fugacity measurements on minerals from mafic rocks that cool at pressures greater than one atmosphere indicate that the Earth's interior is in a more reduced state than are the erupting tholeiitic magmas. This review suggests that specific oxidation–reduction equilibria within magmas and between magma and vapour can account for the characteristic oxygen fugacities of most lavas. A model is presented to illustrate how an initially highly reduced mafic magma may oxidize by degassing of C-species in the conditions of the shallow crust.

SINCE the discovery of the relationship¹ between oxygen fugacity (f_{O_2}) and Fe^{2+}/Fe^{3+} ratios of silicate melts, the characterization of f_{O_2} s of mafic magmas and their mantle source regions has been an important goal in igneous petrology.

Concern for f_{O_2} originated with the recognition that fractional crystallization of basaltic magma under various f_{O_2} s could, at least in principle, account for a wide spectrum of rock compositions². Osborn³, for example, argued that members of cumulate sequences, such as those exemplified by the Skaergaard intrusion, could be related by fractional crystallization of basaltic magma in relatively reducing conditions (the Fenner trend), whereas melt compositions represented by the calc-alkaline sequence basalt–andesite–dacite–rhyolite (the Bowen trend) could be accounted for by fractional crystallization in oxidizing conditions. Thus, the polemic of how basaltic magmas typically differentiate seemed to be resolved. However, as both experiments^{4,5} and data from natural rocks^{6–8} accumulated, it became apparent that mafic magmas are generally characterized by f_{O_2} s which are too low to influence appreciably the course of their differentiation or account for the genesis of calc-alkaline rocks.

Current interest in the oxidation state stems from the strong influence that f_{O_2} exerts on the stability of sulphide–oxide liquids in silicate melts⁹ (and thus the role it has in development of magmatic ore deposits) and on the stabilities of phases rich in the volatile elements, such as C–O–H fluid and graphite. Knowledge of fluid–graphite stability is particularly relevant to our understanding of the mantle. Graphite may strongly enhance electrical conductivities, if it exists as an interconnecting phase on grain boundaries¹⁰; fluid should promote cracking¹¹ and therefore affect mass transport properties and the distributions of rare gases and other incompatible elements.

The various methods available to estimate oxidation states of mafic magmas and their mantle source regions do not lead to a consistent view of the nature of the Earth's interior. On one hand, f_{O_2} s of erupting basaltic lavas, whether determined by direct measurement⁶, deduced from compositions of coexisting gases¹² or inferred from crystallized oxide minerals¹³, are, with few exceptions, similar to or slightly higher than f_{O_2} s of the synthetic equilibrium assemblage quartz–fayalite–magnetite (the QFM buffer) at equivalent temperatures. This similarity has led to the suggestion that oxidation states of lavas are determined by equilibria established among the various Fe-bearing condensed phases within the Earth, and are not significantly influenced by addition or subtraction of volatile-rich matter in the near-surface environment¹⁴. Implicit in this view is that oxidation states of lavas accurately reflect those of their mantle source regions. However, intrinsic f_{O_2} measurements on phases in rocks

that equilibrated under pressures of the shallow crust or upper mantle^{15–18} suggest that the Earth's interior is more reduced than QFM. This has led to an alternative view that reactions involving the vapour phase and loss of volatiles from magmas exert an important influence on oxidation states^{19,20}. Recent attempts to reconcile these apparently inconsistent observations have focused on the possibility of a mantle heterogeneous in terms of its oxidation state^{21,22}.

Using current conceptions of oxidation states of mafic magmas and their upper mantle source regions, I now consider why quasi-QFM conditions characterize lavas and why their oxidation states may, in general, differ from those of magmas and the mantle. It is argued that specific oxidation–reduction (redox) equilibria within melt and between melt and vapour have a critical role, and a model is presented to show how magmas which are initially highly reduced might evolve towards QFM conditions during degassing of C-species at shallow levels in the crust.

Oxygen fugacity estimates

Because the redox reactions for nearly all synthetic and natural mineral assemblages are strongly temperature- and to a lesser extent pressure-dependent, it is not possible to describe oxidation states by citing specific values of f_{O_2} without also specifying T and P . However, the dependencies of most solid redox reactions on T and P are similar. That is, stability curves of assemblages such as QFM and iron–wüstite (IW) possess nearly identical slopes in $1/T$ versus $\log f_{O_2}$ plots (Fig. 1). Experience demonstrates that rocks cool along $1/T$ – $\log f_{O_2}$ paths which are approximately parallel to those of most synthetic buffer assemblages. Therefore, it is convenient to describe natural assemblages by reference to one of the synthetic buffers¹⁹. However, although QFM is used as a convenient reference, it has no *a priori* significance as far as the Earth is concerned because quartz, fayalite, and magnetite are rarely found together.

There are several means of estimating f_{O_2} s of erupting basalts; these are summarized as follows.

(1) f_{O_2} s have been measured directly with electrochemical probes utilizing solid electrolytes¹⁵ and by field gas chromatography²³. For example, *in situ* measurements through the crust of the Makaopuhi lava lake, Hawaii⁶, and of gases emitted during the eruption of Etna²⁴ indicate that these lavas erupted in QFM-type conditions (that is, $QFM \pm 1 \log$ unit in f_{O_2}). The data from Makaopuhi further established that lavas generally cool along T – f_{O_2} paths approximated by QFM until low-temperature deuteric oxidation processes take place.

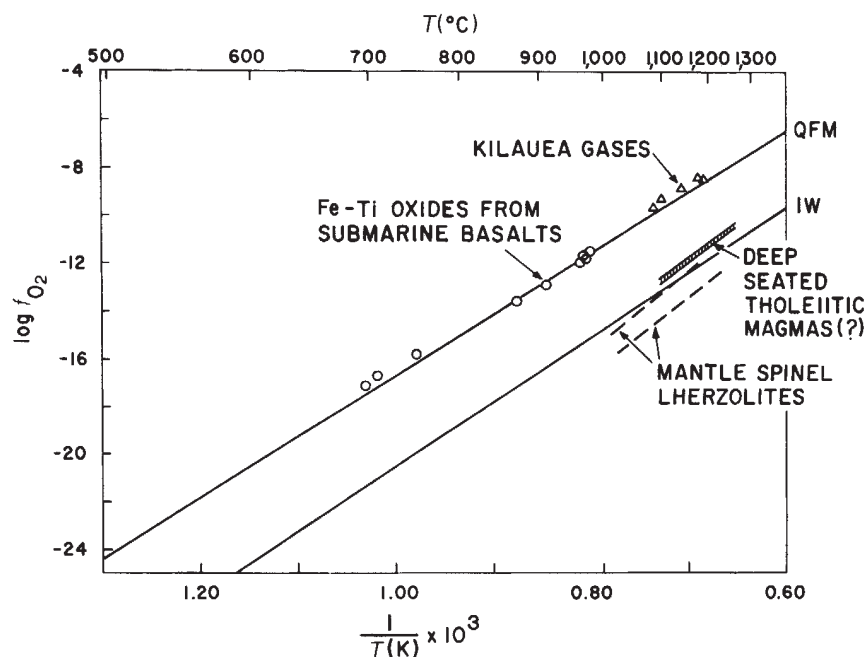


Fig. 1 Comparison of f_{O_2} s of some mafic and ultramafic systems in various conditions. The Kilauea gases are from ref. 25; the Fe-Ti oxide data⁵⁶ are from a thick submarine basalt flow recovered by DSDP Leg 53; the mantle spinels represent the range of intrinsic f_{O_2} measurements of Arculus and Delano¹⁸; and the crustal tholeiitic magma data represent a combination of intrinsic f_{O_2} measurements of Sato¹⁵ on a submarine basalt olivine and Sato and Valenza¹⁶ on minerals from the Lower Zone of the Skaergaard Intrusion and f_{O_2} estimates from C/CO₂ ratios in glass-vapour inclusions in phenocrysts in submarine basalt glasses⁴⁰. The oxygen buffers QFM ($3Fe_2SiO_4 + O_2 \rightleftharpoons 2Fe_3O_4 + 3SiO_2$)⁵⁷ and IW ($Fe + 1/2O_2 \rightleftharpoons FeO$) are shown for reference.

(2) Many volcanic gases, when their compositions are corrected for atmospheric contamination, partial condensation, and/or reactions with collection apparatuses, approximate equilibrium mixtures of gas species at specific magmatic temperatures and f_{O_2} s (ref. 12). Such 'reconstructed' gas compositions from Kilauea²⁵, Surtsey²⁶, Ertä Ale volcano, Ethiopia²⁷, and Merapi volcano, Indonesia²³, indicate that QFM or slightly more oxidized conditions are characteristic for all of the lavas of this diverse suite. In fact, the only gases found to be significantly more reduced than QFM were those associated with the highly alkalic lavas of the Nyiragongo volcano²⁸.

(3) Compositions of coexisting ilmenites and titanomagnetites are related by unique values of temperature and f_{O_2} (ref. 29) and the method has been applied extensively to mafic lavas. For nearly all common basaltic or andesitic bulk compositions, the Fe-Ti oxide minerals do not precipitate until late in the melt crystallization histories. Furthermore, the oxides readily re-equilibrate with falling temperature and therefore typically record equilibria established in the range of 600–1,000 °C in rapidly quenched rocks. In plutonic rocks, low-temperature 'oxidation-exsolution'¹³ of the oxides limits their usefulness for deducing high-temperature conditions. Haggerty's³⁰ compilation of analyses of coexisting oxides demonstrates that basalts and andesites unaffected by deuteric processes invariably exist under QFM-type f_{O_2} s. This conclusion remains unchanged by reformulations of the oxide geobarometer³¹.

(4) Fudali⁸ determined experimentally that the f_{O_2} s required to reproduce the Fe^{2+}/Fe^{3+} ratios observed in natural basaltic to andesitic lavas are at, or slightly above, QFM. Subsequent experimentation has established that melts characterized by such conditions will crystallize minerals having compositions similar to, and in the crystallization sequences of those of, natural assemblages^{32,33}.

Intrinsic f_{O_2} measurements on spinels from mantle lherzolites brought to the surface as xenoliths in alkali basalts¹⁸ imply that oxidation states of some parts of the mantle are near IW (Fig. 1). As pointed out by Virgo *et al.*¹⁷, the validity of 1 atmosphere measurements on minerals cooled at high pressures assumes that pressure does not affect the oxidation states of Fe (or other elements) in minerals. However, even a subtle departure of a mineral's high-pressure stoichiometry from its low-pressure one can translate into large differences between its high- and low-pressure f_{O_2} s of equilibration.

If the above measurements are accepted at face value, other

observations indicate that the mantle must be heterogeneous in terms of oxidation states. First, nearly all mantle xenoliths from alkali basalts contain high-pressure CO₂-rich fluid inclusions³⁴. Such fluids are not stable in a mantle as reduced as IW. Second, relatively oxidizing conditions are indicated by intrinsic f_{O_2} measurements on ilmenite megacrysts from kimberlites³⁵. Third, such conditions are also indicated by compositions of coexisting minerals in some peridotite xenoliths from kimberlites^{21,22,35}. Based on these observations, it has been argued that oxidation states are determined by equilibria involving graphite (or diamond), carbonate and silicates in at least those parts of the mantle sampled by kimberlites. The f_{O_2} s dictated by such assemblages are only slightly below those of QFM at pressures of the xenolith source regions.

One reason why a clear view of the oxidation state of the Earth's interior does not emerge from studies of lavas or mantle-derived rocks may be that f_{O_2} s of magmas are influenced by processes occurring in the crust. This is indicated by certain observations from submarine basalts and layered intrusions. On the one hand, intrinsic f_{O_2} measurements on early-crystallized minerals from stratigraphically low levels of the Skaergaard¹⁶, Bushveld³⁶, and Stillwater³⁶ complexes generally indicate that the f_{O_2} s of the first undifferentiated magmas were near those of the IW buffer. These measurements are thus consistent with those on mantle spinels. In contrast, similar measurements on minerals from generally higher stratigraphical levels indicate more oxidized conditions. Also, magnetites are common in upper horizons of the layered complexes. This requires that the f_{O_2} s of the late stage magmas from which they crystallized were at or above those of the magnetite-wüstite (MW) buffer, which defines the lower limit of magnetite stability. At greater than 1,000 °C, the f_{O_2} s of this buffer are only about 1 log unit below those of QFM. Because the layered complexes cooled under moderate pressures in the shallow crust (< 5 kbar), intrinsic f_{O_2} measurements on constituent minerals are probably not burdened by the uncertainties associated with similar measurements on mantle minerals. In any case, the intrinsic f_{O_2} data together with the occurrences of magnetite indicate that the layered intrusions became more oxidized sometime during their crystallization histories.

In the case of mid-ocean ridge basalts (MORB), Fe-Ti oxides typically have compositions that are similar to those in subaerially erupted tholeiites. For example, the oxides in the thick submarine flows encountered by DSDP Leg 53 demonstrate that

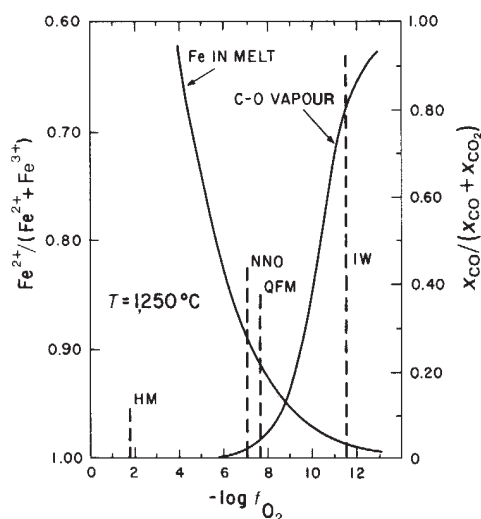


Fig. 2 The oxidation state of Fe in basaltic melt, as computed from the calibration of Sack *et al.*³⁹, and the molecular composition of pure C–O vapour as functions of f_{O_2} . The melt composition is taken as that of DSDP glass sample 3-18-7-1³⁸, which is one of the most primitive rocks recovered from the ocean floor. In all the figures, gas compositions are computed for 1 kbar and 1,250 °C from thermodynamic data compiled by Robie *et al.*⁵⁸. Fugacity coefficients for both pure gases and their mixtures were computed from a Redlich–Kwong equation of state⁴⁴.

these rocks cooled along T – f_{O_2} paths almost identical to QFM (Fig. 1), in agreement with similar data collected for many other submarine rocks³⁷. In addition, the $Fe^{3+}/(Fe^{3+} + Fe^{2+})$ ratios of submarine basalt glasses are typically in the range 0.1–0.2 (ref. 38), which at magmatic temperatures is equivalent to f_{O_2} s near QFM, according to the calibration of Sack *et al.*³⁹.

Other features of the submarine rocks suggest that MORB magmas may have been considerably more reduced before their eruption. Sato¹⁵ reported the intrinsic f_{O_2} of an olivine phenocryst separated from a tholeiitic glass recovered from the submarine environment near Hawaii, and average relative abundances of C and CO₂ in vesicles of glass inclusions in phenocrysts from MORBs, have been estimated⁴⁰ to infer magmatic f_{O_2} s. These data suggest that f_{O_2} s of MORB magmas are near IW. They are thus similar to f_{O_2} estimates of the first magmas of layered intrusions and are clearly inconsistent with f_{O_2} estimates based on Fe–Ti oxide or matrix glass compositions.

Many glass–vapour inclusions in phenocrysts of submarine basalts were entrapped by host phenocrysts in crustal magma chambers before substantial differentiation or degassing^{41–43}. Sato's olivine and the early crystallizing minerals from layered intrusions are also likely to have equilibrated with essentially undegassed magma. In contrast, by the time that magma erupts and quenches to glass on the ocean floor, it is almost completely degassed with respect to C-species. By the time Fe–Ti oxides appear (typically after 70–75% crystallization for tholeiitic compositions) a substantial proportion of the H-species have probably disappeared as well. Thus, undegassed magmas seem to be considerably more reduced than degassed ones.

Controls on oxygen fugacity

Various degassing mechanisms have been postulated to influence oxidation states of magmas. The specific effects of outgassing depend on the abundances of the volatile elements, the manner in which they are dissolved in melts, their speciation in the vapour, and the mass balance relations between the two phases. If magmas are generally open to volatiles—either because they degas or react with assimilated meteoric fluid—then we still need to explain why QFM conditions are so typical of eruption. The answer must lie in the nature of the redox-controlling reactions. Those that operate both in the melt and vapour are now considered.

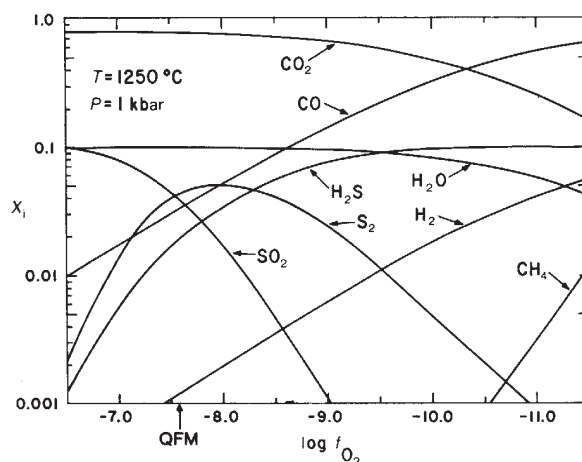


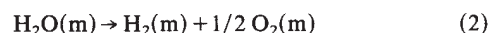
Fig. 3 Molecular composition of magmatic gas as a function of f_{O_2} . The gas composition is such that $X_{CO_2} + X_{CO} + X_{CH_4} = 0.8$, $X_{H_2O} + X_{H_2} = 0.1$, and $X_{H_2S} + X_{S_2} + X_{SO_2} = 0.1$. Note that such a gas does not remain constant in terms of its C/H or S/C as f_{O_2} varies. The ratio of CO₂ to CO is not strongly influenced by the presence of H and S species for C-rich gases in the specified conditions. The C-rich composition arbitrarily chosen here should approximate that of early-formed vapour associated with tholeiitic magmas in the crust (see text).

The equilibrium between Fe^{2+} and Fe^{3+} in magmas is approximated by the relationship⁸



This relationship has been calibrated recently by Sack *et al.*³⁹ for a wide range of silicate melt compositions. The relationship between f_{O_2} and the relative concentrations of Fe^{2+} and Fe^{3+} dissolved in a primitive MORB composition is shown in Fig. 2.

Consider a mechanism which tends to increase a magma's oxidation state. One such mechanism might be the continuous outward diffusion of H from magma, resulting in continuous dissociation of H₂O and increase in O concentration in the residual melt (m):



This process was first proposed by Sato and Wright⁶ to account for certain oxidation features observed in the crusts of Hawaiian lava lakes, and it is supported by observations of H₂ discharge from active volcanic areas¹⁶.

As can be seen in Fig. 2, increasing the f_{O_2} of melt by nearly 4 orders of magnitude from IW to QFM changes its $Fe^{2+}/(Fe^{2+} + Fe^{3+})$ ratio from 0.986 to 0.912. Such an oxidation requires an addition of 0.0012 moles of O₂ per 100 moles of melt, which is equivalent to the dissociation of 0.07 wt % H₂O for the particular bulk composition used in Fig. 2. However, a further increase in f_{O_2} of only 2 orders of magnitude (from $10^{-7.65}$ to $10^{-5.65}$) requires dissociation of an additional 0.12 wt % H₂O.

As f_{O_2} increases, reaction (1) exerts a progressively stronger buffering influence on processes that tend to change f_{O_2} . In relatively reduced conditions, such as near the IW buffer, the relative concentrations of Fe^{2+} and Fe^{3+} in magmas are relatively insensitive to even orders of magnitude changes in f_{O_2} . An important corollary is that equilibria among the condensed Fe-bearing phases (such as, olivine–liquid) are not strongly dependent on f_{O_2} in relatively reduced systems but become more so in those that are as oxidized as QFM or greater. Additionally, this analysis should emphasize that the H diffusion mechanism of Sato and Wright⁶ is potentially a very effective means of modifying oxidation states of magmas.

Speciation within the vapour phase is strongly dependent on f_{O_2} (ref. 44). This dependency is shown in Fig. 3 for a complex C–O–H–S magmatic gas of relatively C-rich bulk composition.

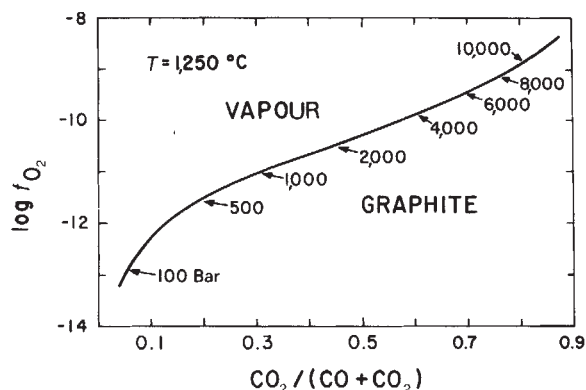
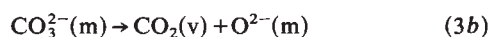
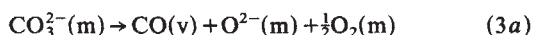


Fig. 4 Compositions of pure C-O vapour in equilibrium with graphite as a function of f_{O_2} and total pressure at 1,250 °C. The equilibrium is $2CO \rightleftharpoons C + CO_2$. Note that the stability field of graphite limits CO-rich compositions to low-pressure environments.

Although there is no unique magmatic gas composition, the C-rich one chosen in Fig. 3 should be representative of the first vapour to separate from tholeiitic melt in the shallow crust. This is anticipated because C is known to be much less soluble than H or S in melts at low pressures. It is also consistent with the fact that CO_2 is the dominant gas species associated with submarine basalts⁴⁵. As Fig. 3 shows, the oxidized species (CO_2 , H_2O , SO_2) vastly dominate at f_{O_2} s above approximately QFM, but the reduced species (CO , H_2S , H_2) are important at f_{O_2} s near IW. This general type of behaviour extends throughout the entire C-O-H-S system¹². It also characterizes the simpler C-O system, which is included on Fig. 2. Pure CO- CO_2 gases will be used as models for magmatic ones because: (1) the geological conditions for degassing in the model that follows are such that C-rich C-O-H-S gases are anticipated; and (2) there is essentially no difference in the relative amounts of CO_2 and CO in the C-rich complex gases and the more simple pure C-O ones.

The Raman and IR spectra of synthetic C-rich silicate glasses suggest that C dissolves in magmas as CO_3^{2-} (refs. 46, 47). Furthermore, CO and CO_2 both seem to dissolve by this mechanism⁴⁷, that is, the solubility relations are essentially independent of the relative fugacities of CO and CO_2 or the oxidation state of the system. Therefore, assuming that the conclusions of these vibrational spectra studies are correct and that other solution mechanisms are of negligible importance, there are two specific degassing reactions that can be written for the carbon species:



Reaction (3a) involving production of CO vapour (v) results in an increased O content of the residual melt (m) and thus in an increase in its oxidation state through reaction (1), whereas reaction (3b), in which CO_2 is produced, obviously involves no such change in oxidation state.

Consider a hypothetical reduced magma that is initially supersaturated in C and which degases in conditions of the shallow crust by a process involving a continuous and infinitesimal exsolution of the vapour components from the magma—by a process that will be termed perfect fractional vaporization to emphasize the fact that it is analogous to the process of perfect fractional crystallization. The first vapour to exsolve through reactions (3a) and (3b) will be relatively CO-rich (Fig. 2). Consequently, the melt will oxidize through reaction (1), and succeeding vapour fractions will become progressively more CO_2 -rich until all available C is exhausted (until the system is no longer supersaturated). During this process f_{O_2} will increase

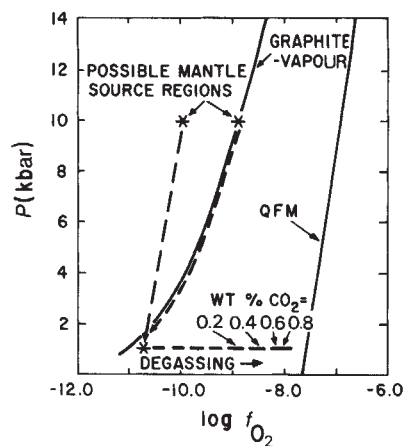


Fig. 5 Possible P - f_{O_2} paths of submarine basalt magmas. Degassing in a shallow crustal chamber causes oxidation of the magma. The increase in f_{O_2} of a primitive tholeiitic melt (of composition similar to that for Fig. 2) from degassing by perfect fractional vaporization is shown for loss of various amounts of C expressed as wt% CO_2 . The graphite-vapour curve is the same as that shown in Fig. 4.

at a progressively slower rate as reaction (3b) gradually replaces (3a) as the important degassing one and as reaction (1) consumes progressively more O. In other words, both the vapour and the melt begin exerting stronger buffering influences on f_{O_2} as f_{O_2} increases (Fig. 2).

An important limitation of this mechanism for generating more oxidized conditions in the real Earth is that it can operate only at low pressures. This is because CO-rich vapours are not stable at high pressures as a result of the limitation on vapour compositions imposed by the field of graphite stability, as shown in Fig. 4. It can be seen, for example, that magmatic vapours consisting of more than 50% CO can only exist at pressures of <3 kbar. Thus, it is necessary to envisage emplacement of a supersaturated magma in a shallow crustal reservoir before substantial degassing occurs.

This limitation means that the analogy between the degassing process and fractional crystallization is not precisely accurate. It is offered, however, as a means of conceptualizing the process and allowing calculation of the extent to which the melt will be oxidized. The analogy arises if it is assumed that degassing occurs sufficiently slowly for the residual magma to exsolve succeeding vapours of progressively more CO_2 -rich compositions. The degassing reactions are not reversible because they are driven by supersaturation of the magma in C-species, which results in decompression and/or crystallization, just as fractional crystallization is driven by supersaturation in 'crystal' components due to cooling. Note that the 'system' in both of these processes is regarded as the melt only. Thus, the bulk composition of the system changes as vapour exsolves, regardless of whether the vapour remains as bubbles suspended in the melt or escapes from the magma chamber. If, however, vapour and melt remain in contact so that they are allowed to approach a final equilibrium state, that final state is the same regardless of how rapidly degassing proceeds or what path it assumes.

Reactions (3a) and (3b) along with the relations plotted in Fig. 2 allow the extent of oxidation attendant on degassing to be computed according to the perfect fractional vaporization mechanism. From assumed initial values of f_{O_2} and C contents of a given melt composition, this can be done by determining from the calibration of Sack *et al.*³⁹ the amount of O required to change the f_{O_2} of the melt by a chosen small increment, by computing the amount of CO_3^{2-} that must be degassed according to reactions (3a) and (3b) to supply the required O, and then by repeating the entire calculation for subsequent f_{O_2} increments

until all C is consumed by degassing. A geological situation in which this degassing mechanism might reasonably be approached remains to be considered.

Application to submarine basalts

Consider a primitive submarine tholeiitic magma emplaced in a shallow chamber at 1 kbar beneath a spreading centre. The presence of such chambers at pressures of <3 kbar in the oceanic crust is indicated by geophysical data⁴⁸. Magma temperature is assumed to be 1,250 °C and its f_{O_2} to be consistent with the data for undegassed submarine basalts and other crustal melts (Fig. 1), is taken to be three orders of magnitude below that of QFM in equivalent conditions ($10^{-10.6}$ or approximately 1 log unit above IW) (Fig. 5). The manner in which the magma arrived at these conditions is unimportant. For example, it may have originated from a highly reduced mantle source at 10–12 kbar and followed a P – f_{O_2} path to the crust approximately parallel to QFM. This simply implies that f_{O_2} s are dictated by reactions involving the major silicate crystalline phases and that graphite rather than fluid is the major C-bearing phase in the sub-oceanic mantle. Such a highly reduced mantle is also consistent with some of the intrinsic f_{O_2} measurements on mantle rocks, as noted previously. Alternatively, it may have evolved along a P – f_{O_2} path defined by equilibrium between graphite and vapour, as suggested by Sato¹⁹, in which case the source region would have been more oxidized (Fig. 5). In either case, it is important that the magma arrive in the shallow crust in a supersaturated state.

The solubility of C in tholeiitic melt at high pressure is poorly known. Based on Spera and Bergman's⁴⁹ analysis of available experimental data and following later work⁴⁶, tholeiitic magmas generated at 10–12 kbar might be expected to contain perhaps 500–1,500 p.p.m. dissolved C. However, Pineau and Javoy⁵⁰ have estimated from isotopic data that C abundances in undegassed MORB magmas are typically in the range 2,000–10,000 p.p.m.

Fortunately, precise knowledge of initial C abundances is not particularly critical to the oxidation model because of the shapes of the curves in Fig. 2. Thus, a magma that originally contains only 0.20 wt% C as CO₂ (546 p.p.m. C) and loses all C by degassing in the conditions specified above will oxidize such that f_{O_2} increases from $10^{-10.6}$ to $10^{-8.9}$. One which originally contained 0.40 wt% CO₂ would oxidize from an f_{O_2} of $10^{-10.6}$ to $10^{-8.5}$. Further oxidation of this magma to an f_{O_2} of $10^{-8.0}$ would require an additional 0.39 wt% CO₂ (see Fig. 5).

Solubilities of S and H in mafic melts are large compared with that of C. Hence, a large proportion of the original amounts

of S and H appear to remain in solution in glasses formed by quenching at pressures of the ocean floor⁵¹, while the associated gases are CO₂-rich^{45,50}. In contrast, S and H almost completely degas from lavas that erupt at 1 atmosphere, and even in submarine basalts H should exsolve from lavas that crystallize anhydrous phases rather than quench to glasses. In the case of H, degassing should not significantly influence magmatic oxidation states because H dissolves in melts as H₂O and OH (refs 52, 53) and H₂O is the dominant H species in moderately oxidized C–O–H–S gases⁴⁴. (This neglects the H diffusion mechanism of Sato and Wright⁶, which may be important but which is not strictly a degassing process.) In contrast, the specific effect of degassing of S-species on oxidation state is not easily predicted. This is because the manner in which S dissolves in melts is itself strongly dependent on f_{O_2} (refs 54, 55), solubility mechanisms are probably not completely understood (for example, H₂S may be an important dissolved species in some melts) and speciation of S in C–O–H–S gases is more complicated than for C or H. The gas composition plotted on Fig. 3 was chosen in part to illustrate this problem. Speculation on the influence of S degassing is, therefore, not appropriate at this point.

Conclusion

Figure 5 shows that the f_{O_2} s of tholeiitic magmas can be elevated to those approaching QFM only if CO₂ contents are of the order of 1%. Whether or not such compositions exist in the Earth, or the kinetics of vapour exsolution are such that supersaturated melts can reach shallow crustal levels, should be matters of future debate. This model simply shows that by choice of a reasonable and not unlikely set of geological circumstances, degassing of C-species alone may influence magmatic f_{O_2} s if primary magmas are as reduced as many of the intrinsic f_{O_2} measurements would suggest. If this is, in fact, the case, then it is not possible to deduce oxidation states of magmas or their mantle source regions from compositions of volcanic gases or coexisting Fe–Ti oxides in lavas. Rather than being indicative of conditions in the Earth's interior, that lavas generally erupt in approximately QFM-type conditions may reflect the strong buffering influence that equilibrium between Fe²⁺ and Fe³⁺ in silicate melts exerts on f_{O_2} at f_{O_2} s near QFM.

I thank all those who have discussed, criticized or commented on versions of this manuscript. They include I. S. McCallum, B. W. Evans, M. Sato, T. L. Grove, and A. J. Irving. The support of NSF grant OCE 8110149 is gratefully acknowledged.

- Kennedy, G. C. *Am. J. Sci.* **246**, 529–549 (1948).
- Kennedy, G. C. *Geol. Soc. Am. Spec. Pap.* **62**, 483–533 (1955).
- Osborn, E. F. *Am. J. Sci.* **257**, 609–647 (1959).
- Roeder, P. L. & Osborn, E. F. *Am. J. Sci.* **264**, 428–480 (1966).
- Presnall, D. C. *Am. J. Sci.* **264**, 753–809 (1966).
- Sato, M. & Wright, T. L. *Science* **153**, 1103–1105 (1966).
- Anderson, A. T. *Am. J. Sci.* **266**, 704–727 (1968).
- Fudali, R. F. *Geochim. cosmochim. Acta* **29**, 1063–1075 (1965).
- Haughton, D. R., Roeder, P. L. & Skinner, B. J. *Econ. Geol.* **69**, 451–467 (1974).
- Duba, A. G. & Shankland, T. J. *Geophys. Res. Lett.* **9**, 1271–1274 (1982).
- Anderson, O. L. & Grew, P. C. *Rev. Geophys. Space Phys.* **15**, 77–104, (1977).
- Gerlach, T. M. & Nordlie, B. E. *Am. J. Sci.* **275**, 377–394 (1975).
- Haggerty, S. E. *Miner. Soc. Am. Short Course Not.* **3**, 1–100 (1976).
- Haggerty, S. E. *Geophys. Res. Lett.* **5**, 443–446 (1978).
- Sato, M. *Mem. geol. Soc. Am.* **135**, 289–307 (1972).
- Sato, M. & Valenza, M. M. *J. Sci.* **280**–A, 134–158 (1980).
- Virgo, D., Rosenhauer, M. A. & Huggins, F. E. *Yb Carnegie Instn Washington* **75**, 720–735 (1976).
- Arculus, R. J. & Delano, J. W. *Geochim. cosmochim. Acta* **45**, 889–913 (1981).
- Sato, M. *Geophys. Res. Lett.* **5**, 447–449 (1978).
- Sato, M. *Proc. 10th Lunar planet. Sci. Conf.* **311**–325 (1979).
- Eggler, D. H. *Geophys. Res. Lett.* **10**, 365–368 (1983).
- Haggerty, S. E. & Tompkins, L. A. *Nature* **303**, 295–300 (1983).
- LeGuern, F., Gerlach, T. M. & Nohl, A. J. *Volcan. Geotherm. Res.* **14**, 223–245 (1982).
- Sato, M. & Moore, J. G. *Phil. Trans. R. Soc. A* **274**, 137–146 (1973).
- Gerlach, T. M. *J. Volcan. Geotherm. Res.* **7**, 295–317 (1980).
- Gerlach, T. M. *J. Volcan. Geotherm. Res.* **8**, 191–198 (1980).
- Gerlach, T. M. *J. Volcan. Geotherm. Res.* **7**, 415–441 (1980).
- Gerlach, T. M. *J. Volcan. Geotherm. Res.* **8**, 177–189 (1980).
- Buddington, A. F. & Lindsley, D. H. *J. Petrol.* **5**, 310–357 (1964).
- Haggerty, S. E. *Miner. Soc. Am. Short Course Not.* **3**, 101–300 (1976).
- Stormer, J. C. Jr *Am. Miner.* **68**, 586–594 (1983).
- Duke, J. M. *J. Geol.* **82**, 524–532 (1974).
- Hill, R. & Roeder, P. J. *Geol.* **82**, 709–719 (1974).
- Bergman, S. C. & Debessy, J. *Contr. Miner. Petrol.* **85**, 1–13 (1984).
- Arculus, R. J., Dawson, J. B., Mitchell, R. H., Gust, D. A. & Holmes, R. D. *Contr. Miner. Petrol.* **85**, 85–94 (1984).
- Elliot, W. C., Grandstaff, D. E., Ulmer, G. C., Buntin, T. & Gold, D. P. *Econ. Geol.* **77**, 1493–1510 (1982).
- Mazzullo, L. J. & Bence, A. E. *J. geophys. Res.* **81**, 4327–4351 (1976).
- Frey, F. A., Bryan, W. B. & Thompson, G. J. *geophys. Res.* **79**, 5507–5527 (1974).
- Sack, R. O., Carmichael, I. S. E., Rivers, M. & Ghiorsio, M. S. *Contr. Miner. Petrol.* **75**, 369–376 (1980).
- Mathez, E. A. & Delaney, J. R. *Earth planet. Sci. Lett.* **56**, 217–232 (1981).
- Watson, E. B. *J. Volcan. Geotherm. Res.* **1**, 73–84 (1976).
- Delaney, J. R., Muenow, D. W. & Graham, D. G. *Geochim. cosmochim. Acta* **42**, 581–594 (1978).
- Harris, D. M. *J. Geol.* **89**, 689–701 (1981).
- Holloway, J. R. *Thermodynamics of Minerals and Melts* (eds Newton, R. C., Navrotsky, A. & Wood, B. J.) 273–293 (Springer, New York, 1981).
- Moore, J. G., Butchelder, J. N. & Cunningham, C. G. *J. Volcan. Geotherm. Res.* **2**, 309–327 (1977).
- Rai, C. S., Sharma, S. K., Muenow, D. W., Matson, D. W. & Byers, C. D. *Geochim. cosmochim. Acta* **47**, 953–958 (1983).
- Eggler, D. H., Mysen, B. O., Hoering, T. C. & Holloway, J. R. *Earth planet. Sci. Lett.* **43**, 321–330 (1979).
- Lewis, B. T. R. *J. geophys. Res.* **88**, 3348–3354 (1983).
- Spera, F. J. & Bergman, S. C. *Contr. Miner. Petrol.* **74**, 55–66 (1980).
- Pineau, F. & Javoy, M. *Earth planet. Sci. Lett.* **62**, 239–257 (1983).
- Moore, J. G. & Schilling, J. G. *Contr. Miner. Petrol.* **41**, 105–118 (1973).
- Burnham, C. W. *Geochim. cosmochim. Acta* **39**, 1077–1084 (1975).
- Stolper, E. *Geochim. cosmochim. Acta* **46**, 2609–2620 (1982).
- Katsura, T. & Nagashima, S. *Geochim. cosmochim. Acta* **38**, 517–531 (1974).
- Sakai, H., Casadevall, T. J. & Moore, J. G. *Geochim. cosmochim. Acta* **46**, 729–738 (1982).
- Gitlin, E. C. thesis, Univ. Washington (1981).
- Hewitt, D. A. *Am. J. Sci.* **278**, 715–724 (1978).
- Robie, R. A., Hemingway, B. S. & Fisher, J. R. *U.S. geol. Surv. Bull.* **1452** (1978).

3-Å resolution structure of a protein with histone-like properties in prokaryotes

Isao Tanaka*, Krzysztof Appelt, Jan Dijk, Stephen W. White* & Keith S. Wilson*

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33 (Dahlem), FRG

The 3-Å structure of DNA-binding protein II, which exhibits histone-like properties in bacteria, has been determined. The molecule is dimeric and appears to bind to the phosphate backbone of DNA through two symmetry-related arms. A mechanism by which the protein induces DNA supercoiling is proposed.

It is recognized that the enormously long DNA molecules of the genome must be compactly folded in a manner which allows their accommodation within the cell without hindering their accessibility. In eukaryotes, the DNA interacts with the histone proteins¹ to form a tightly packed superstructure known as chromatin. At the first level of organization, the histones form a so-called core particle around which the DNA is coiled. The core particles pack together at higher levels of organization, ultimately leading to the formation of the chromosome². The structure of the complete core particle is currently under investigation³. However, no structure is available for any of the histone proteins, although crystals of the histone core have been isolated⁴.

In prokaryotes, although the genome is much smaller, it is clear that in both eukaryotes and prokaryotes there is DNA packaging. Following the gentle lysis of *Escherichia coli* cells, chromosomes can be isolated in a folded and supercoiled conformation⁵ and these have been visualized in the electron microscope⁶. In addition, several DNA-binding proteins have been identified⁷ for which a histone-like role has been proposed. One of these proteins, which according to a recent nomenclature system⁷ we shall refer to as DNA-binding protein II, exhibits clear histone-like properties. This small basic protein of molecular weight 9,500 has been shown to induce the formation of bead-like structures along double-stranded DNA in a manner similar to the histones⁸. It binds nonspecifically to both single-stranded and double-stranded DNA as well as to RNA⁹, and has been found in a variety of organisms including archaeobacteria^{10–16}. The available sequence information shows it to be highly conserved^{13,16–19}. In *E. coli* there are two subspecies of the protein which have been referred to variously as NS1 and NS2 (ref. 20) or HU α and HU β (ref. 21). However, in all the other organisms studied, only one species has been detected. An analysis of the proteins associated with *E. coli* chromosomes revealed two major species with approximate molecular weights of 17,000 and 9,000 (ref. 22). The latter displayed electrophoretic properties similar to those of DNA-binding protein II and it was subsequently confirmed^{23,24} that, *in vivo*, the protein is associated with the genomic DNA. It has been estimated⁷ that there is sufficient DNA-binding protein II to cover approximately 10% of the chromosomal DNA. Therefore, the 17,000 molecular weight protein (which is probably protein H1)²⁵ and possibly others²⁵ are likely to be necessary for the formation of the complete chromosome.

We have previously reported the isolation and crystallization of DNA-binding protein II from *Bacillus stearothermophilus*¹⁵ and a low resolution analysis of the structure²⁶. We present here the molecular structure of the protein at a resolution of 3 Å.

A mode of DNA binding can be readily deduced from the structure, which is quite different from that proposed for the regulatory DNA-binding proteins that have been studied crystallographically^{27–29}. In addition, the general architecture of the molecule suggests how it might self-associate to form a chromatin-like core particle. Although DNA-binding protein II has no obvious sequence homology to any of the histone proteins, our results may well prove to be relevant to the eukaryotic system.

Structure determination

The crystallization conditions, heavy atom derivative preparation and method of data collection have been described previously^{15,26}. The crystals are in the space group P2₁, with cell dimensions $a = 65.5$ Å, $b = 37.3$ Å, $c = 65.5$ Å and $\beta = 114.5^\circ$. The 6-Å electron density map calculated with single isomorphous phases ($K_3UO_2F_6$) revealed three dimers of DNA-binding protein II in the unit cell located on independent crystallographic 2-fold rotation axes. In the derivative, there is one uranyl-binding site per asymmetric unit located between two of the dimers.

Data were collected to 3 Å for both the native and derivative crystals and, after refinement of the heavy atom parameters, single isomorphous phases were calculated. The resulting 3-Å map showed that each of the three dimers in the unit cell was essentially identical at this resolution (Fig. 1). Therefore, it was possible to average the three electron densities to improve the image of the molecule. As the three dimers are on crystallographic 2-fold axes, there are only two parameters which relate any pair—a shift parameter along the y -axis and a rotational parameter around this axis. These were refined to obtain a maximum overlap between the three electron densities. An averaged electron density was then calculated and placed at each of the three original locations in the unit cell. The new 'averaged' unit cell was Fourier transformed and the resulting phases were used with the observed amplitudes to calculate a new map (Fig. 1). Each of the three new images was of comparable quality and the path of the polypeptide backbone could be traced, apart from the extremity of an arm which extends away from the molecule and for which the density is very weak.

A model was built on an Evans and Sutherland PS300 picture system using the FRODO package of programs. It was possible to fit all the amino acid sequence¹⁹ except for residues 59–70, which correspond to the poorly defined arm region. To check that the three independent dimers have no important conformational differences, the final model was fitted into the non-averaged map. No significant differences were apparent at this resolution.

Structure of the monomer

The monomer structure (Fig. 2) has two distinct halves. The N-terminal half consists of two α -helices ($\alpha 1$, residues 3–13; $\alpha 2$, residues 21–37) which are connected by a broad turn

* Present addresses: Department of Applied Chemistry, Faculty of Engineering, Nagoya University, Chikusa, Nagoya, 464, Japan (I.T.); Biology Department, Brookhaven National Laboratory, Upton, Long Island, New York 11973, USA (S.W.W.); Department of Physics, University of York, Heslington YO1 5DD, UK (K.S.W.).

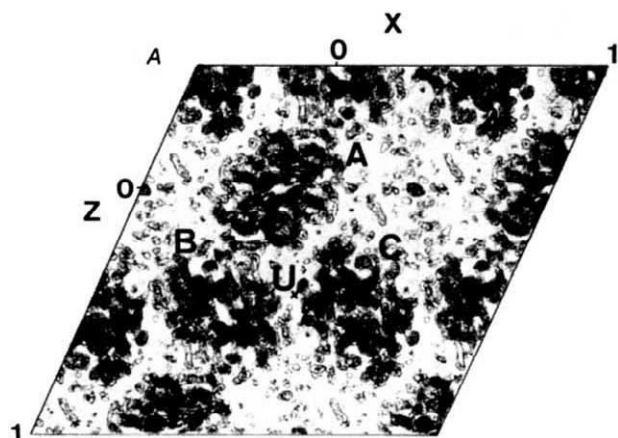
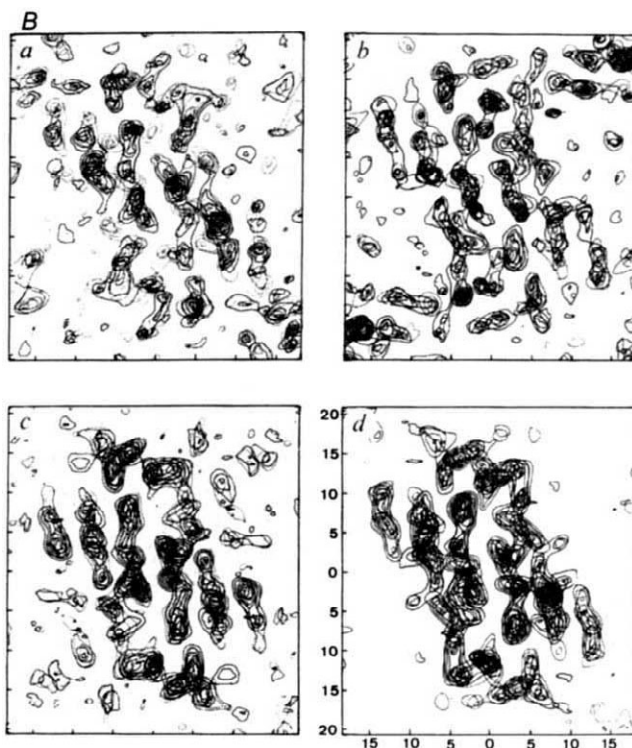


Fig. 1 Electron density maps of DNA-binding protein II. **A**, 3-Å single isomorphous replacement map showing the three independent dimers (A, B, C) in the unit cell. The complete cell is superimposed in the *y* direction (*b*-axis unique). The uranyl heavy atom site (U) is located between molecules B and C and refined to an *R*-factor of 48.6% (acentric). The average figure of merit is 66%. **B**, The density averaging for DNA-binding protein II. *a*, *b* and *c* are the three independent electron densities and *d* is the average. The map sections show a 6-Å slab of density in the region of the β -pleated sheet and the flanking C-terminal helices.



(residues 14–20) to create a V-shaped supersecondary structure. The contact between the helices at the base of the V is mediated by two conserved alanine residues at positions 11 and 21 (Fig. 3). The C-terminal half consists mainly of a three-stranded antiparallel β -pleated sheet (strand 1, residues 40–44; strand 2, residues 48–51; strand 3, residues 78–83) which spans the top of the V formed by the two helices. Strands 1 and 2 are linked by a sharp turn which is highly conserved, particularly the Gly-Phe-Gly sequence between residues 46 and 48 (Fig. 3). However, strands 2 and 3 are connected through an extended arm region. The molecule ends in a short α -helix ($\alpha 3$, residues 84–89). The two halves of the structure are connected by a β -turn between $\alpha 2$ and strand 1 which has a highly conserved glycine (residue 39, see Fig. 3).

The segment of the arm proximal to the main body of the protein has a clear two-strand antiparallel β conformation (a so-called β -ribbon) but we are unable to describe the structure of the distal segment. However, we suggest a possible conformation based on an inspection of the sequence and its conservation in this region (see Fig. 3). The totally conserved proline at

position 63 is ideally placed to occupy position 2 of a β -turn at the end of the arm. The 'outgoing' strand would then contain 11 residues and the 'return' strand 13. By incorporating two β -bulges or their equivalent, the β -ribbon structure could be continued to the end of the arm. The highly conserved proline at position 72 may be the site of one of the β -bulges. This proposed arm conformation has certain attractive features in connection with the possible protein–DNA-binding mechanism which is discussed later.

Structure of the dimer

Two monomers literally wrap around each other to produce a novel dimeric structure (Fig. 4). The base of the molecule is formed by the two V-shaped helical supersecondary structures which pack such that the two $\alpha 2$ helices are in contact roughly at right angles and the two $\alpha 1$ helices are on either side. The conserved alanine residue at position 24 (Fig. 3) marks the $\alpha 2$ – $\alpha 2$ contact. This base is covered by the two three-stranded sheets which, due to the distance between the two strands 3 and

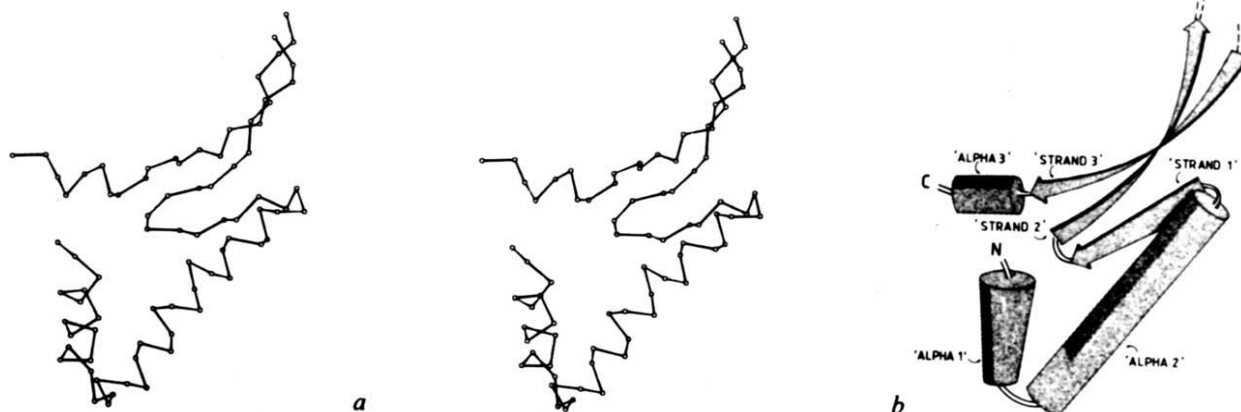


Fig. 2 The structure of the monomer of DNA-binding protein II. **a**, A stereo view of the $C\alpha$ backbone. The residues in the distal region of the arm between strands 2 and 3 (59–70) are not visible in the electron density map and are not shown. **b**, A schematic representation of the secondary structure.

RESIDUE NUMBER	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
B. st.	<u>M</u>	<u>N</u>	<u>K</u>	<u>T</u>	<u>E</u>	<u>L</u>	<u>I</u>	<u>N</u>	<u>A</u>	<u>V</u>	<u>A</u>	<u>E</u>	<u>T</u>	<u>S</u>	<u>G</u>	<u>L</u>	<u>S</u>	<u>K</u>	<u>K</u>	<u>D</u>	<u>A</u>	<u>T</u>	<u>K</u>	<u>A</u>	<u>V</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>F</u>	<u>D</u>	
E. c.1	<u>M</u>	<u>N</u>	<u>K</u>	<u>S</u>	<u>Q</u>	<u>L</u>	<u>I</u>	<u>D</u>	<u>K</u>	<u>I</u>	<u>A</u>	<u>A</u>	<u>G</u>	<u>A</u>	<u>D</u>	<u>I</u>	<u>S</u>	<u>K</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>G</u>	<u>R</u>	<u>A</u>	<u>L</u>	<u>D</u>	<u>A</u>	<u>I</u>	<u>I</u>	<u>A</u>	
E. c.2	<u>M</u>	<u>N</u>	<u>K</u>	<u>T</u>	<u>Q</u>	<u>L</u>	<u>I</u>	<u>D</u>	<u>V</u>	<u>I</u>	<u>A</u>	<u>E</u>	<u>K</u>	<u>A</u>	<u>E</u>	<u>L</u>	<u>S</u>	<u>K</u>	<u>T</u>	<u>Q</u>	<u>A</u>	<u>K</u>	<u>A</u>	<u>A</u>	<u>L</u>	<u>E</u>	<u>S</u>	<u>T</u>	<u>L</u>	<u>A</u>	
C. p.	<u>M</u>	<u>N</u>	<u>K</u>	<u>A</u>	<u>E</u>	<u>L</u>	<u>I</u>	<u>T</u>	<u>S</u>	<u>M</u>	<u>A</u>	<u>E</u>	<u>K</u>	<u>S</u>	<u>K</u>	<u>L</u>	<u>T</u>	<u>K</u>	<u>K</u>	<u>D</u>	<u>A</u>	<u>E</u>	<u>L</u>	<u>A</u>	<u>L</u>	<u>K</u>	<u>A</u>	<u>L</u>	<u>I</u>	<u>E</u>	
R. m.	<u>M</u>	<u>N</u>	<u>K</u>	<u>N</u>	<u>E</u>	<u>L</u>	<u>V</u>	<u>A</u>	<u>A</u>	<u>V</u>	<u>A</u>	<u>D</u>	<u>K</u>	<u>A</u>	<u>G</u>	<u>L</u>	<u>S</u>	<u>K</u>	<u>A</u>	<u>D</u>	<u>A</u>	<u>S</u>	<u>S</u>	<u>A</u>	<u>V</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>F</u>	<u>E</u>	
P. a.	<u>M</u>	<u>N</u>	<u>K</u>	<u>S</u>	<u>Q</u>	<u>L</u>	<u>I</u>	<u>D</u>	<u>A</u>	<u>I</u>	<u>A</u>	<u>A</u>	<u>S</u>	<u>A</u>	-	-	-	<u>K</u>	<u>A</u>	<u>V</u>	<u>A</u>	<u>G</u>	<u>K</u>	<u>A</u>	<u>L</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>I</u>	<u>E</u>	
T. a.	M	V	G	I	S	<u>E</u>	<u>L</u>	S	K	E	<u>V</u>	<u>A</u>	K	K	A	N	T	T	Q	<u>K</u>	<u>V</u>	<u>A</u>	R	T	V	I	K	S	F	L	<u>D</u>
SECONDARY STRUCTURE	----- ALPHA 1 -----													*	TURN						*	----- ALPHA 2 -----									

RESIDUE NUMBER	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
B. st.	<u>S</u>	<u>I</u>	<u>T</u>	<u>E</u>	<u>A</u>	<u>L</u>	<u>R</u>	<u>K</u>	<u>G</u>	<u>D</u>	<u>K</u>	<u>V</u>	<u>Q</u>	<u>L</u>	<u>I</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>N</u>	<u>F</u>	<u>E</u>	<u>V</u>	<u>R</u>	<u>E</u>	<u>R</u>	<u>A</u>	<u>A</u>	<u>R</u>	<u>K</u>	<u>G</u>	
E. c.1	<u>S</u>	<u>V</u>	<u>T</u>	<u>E</u>	<u>S</u>	<u>L</u>	<u>K</u>	<u>E</u>	<u>G</u>	<u>D</u>	<u>Q</u>	<u>V</u>	<u>A</u>	<u>L</u>	<u>V</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>T</u>	<u>F</u>	<u>A</u>	<u>V</u>	<u>K</u>	<u>E</u>	<u>R</u>	<u>A</u>	<u>A</u>	<u>R</u>	<u>T</u>	<u>G</u>	
E. c.2	<u>A</u>	<u>I</u>	<u>T</u>	<u>E</u>	<u>S</u>	<u>L</u>	<u>K</u>	<u>E</u>	<u>G</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>Q</u>	<u>L</u>	<u>V</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>T</u>	<u>F</u>	<u>K</u>	<u>V</u>	<u>N</u>	<u>H</u>	<u>R</u>	<u>A</u>	<u>E</u>	<u>R</u>	<u>T</u>	<u>G</u>	
C. p.	<u>S</u>	<u>V</u>	<u>E</u>	<u>E</u>	<u>A</u>	<u>L</u>	<u>E</u>	<u>K</u>	<u>G</u>	<u>E</u>	<u>K</u>	<u>V</u>	<u>Q</u>	<u>L</u>	<u>V</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>T</u>	<u>F</u>	<u>E</u>	<u>T</u>	<u>R</u>	<u>E</u>	<u>R</u>	<u>A</u>	<u>A</u>	<u>R</u>	<u>E</u>	<u>G</u>	
R. m.	<u>T</u>	<u>I</u>	<u>Q</u>	<u>G</u>	<u>E</u>	<u>L</u>	<u>K</u>	<u>N</u>	<u>G</u>	<u>G</u>	<u>D</u>	<u>I</u>	<u>R</u>	<u>L</u>	<u>V</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>N</u>	<u>F</u>	<u>S</u>	<u>V</u>	<u>S</u>	<u>R</u>	<u>R</u>	<u>E</u>	<u>A</u>	<u>S</u>	<u>K</u>	<u>G</u>	
P. a.	<u>S</u>	<u>V</u>	<u>T</u>	<u>G</u>	<u>A</u>	<u>L</u>	<u>K</u>	<u>A</u>	<u>G</u>	<u>D</u>	-	-	-	-	<u>V</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>T</u>	<u>F</u>	<u>A</u>	<u>V</u>	<u>K</u>	<u>E</u>	<u>R</u>	<u>A</u>	<u>A</u>	<u>R</u>	<u>T</u>	<u>G</u>	
T. a.	<u>E</u>	<u>I</u>	<u>V</u>	<u>S</u>	<u>E</u>	<u>A</u>	<u>N</u>	<u>G</u>	<u>G</u>	<u>Q</u>	<u>K</u>	<u>I</u>	<u>N</u>	<u>L</u>	<u>A</u>	<u>G</u>	<u>F</u>	<u>G</u>	<u>I</u>	<u>F</u>	<u>E</u>	<u>R</u>	<u>R</u>	<u>T</u>	<u>Q</u>	<u>G</u>	<u>P</u>	<u>R</u>	<u>K</u>	<u>A</u>	
SECONDARY STRUCTURE	----- ALPHA 2 -----						*	β TURN		*	STRAND 1				*	TURN		*	STRAND 2				*	ARM REGION							

RESIDUE NUMBER	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	
B. st.	<u>R</u>	<u>N</u>	<u>P</u>	<u>Q</u>	<u>T</u>	<u>G</u>	<u>E</u>	<u>E</u>	<u>M</u>	<u>E</u>	<u>I</u>	<u>P</u>	<u>A</u>	<u>S</u>	<u>K</u>	<u>V</u>	<u>P</u>	<u>A</u>	<u>F</u>	<u>K</u>	<u>P</u>	<u>G</u>	<u>K</u>	<u>A</u>	<u>L</u>	<u>K</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>K</u>	
E. c.1	<u>R</u>	<u>N</u>	<u>P</u>	<u>Q</u>	<u>T</u>	<u>G</u>	<u>K</u>	<u>E</u>	<u>I</u>	<u>T</u>	<u>I</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>K</u>	<u>V</u>	<u>P</u>	<u>S</u>	<u>F</u>	<u>R</u>	<u>A</u>	<u>G</u>	<u>K</u>	<u>A</u>	<u>L</u>	<u>K</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>N</u>	
E. c.2	<u>R</u>	<u>N</u>	<u>P</u>	<u>Q</u>	<u>T</u>	<u>G</u>	<u>K</u>	<u>E</u>	<u>I</u>	<u>K</u>	<u>I</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>N</u>	<u>V</u>	<u>P</u>	<u>A</u>	<u>F</u>	<u>V</u>	<u>S</u>	<u>G</u>	<u>K</u>	<u>A</u>	<u>L</u>	<u>K</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>K</u>	
C. p.	<u>R</u>	<u>N</u>	<u>P</u>	<u>R</u>	<u>T</u>	<u>K</u>	<u>E</u>	<u>V</u>	<u>I</u>	<u>N</u>	<u>I</u>	<u>P</u>	<u>A</u>	<u>T</u>	<u>T</u>	<u>V</u>	<u>P</u>	<u>V</u>	<u>F</u>	<u>K</u>	<u>A</u>	<u>G</u>	<u>K</u>	<u>E</u>	<u>F</u>	<u>K</u>	<u>D</u>	<u>K</u>	<u>V</u>	<u>N</u>	<u>K</u>
R. m.	<u>R</u>	<u>N</u>	<u>P</u>	<u>S</u>	<u>T</u>	<u>G</u>	<u>A</u>	<u>E</u>	<u>V</u>	<u>D</u>	<u>I</u>	<u>P</u>	<u>A</u>	<u>R</u>	<u>N</u>	<u>V</u>	<u>P</u>	<u>K</u>	<u>F</u>	<u>T</u>	<u>A</u>	<u>G</u>	<u>K</u>	<u>G</u>	<u>L</u>	<u>K</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>N</u>	
P. a.	<u>R</u>	<u>N</u>	<u>P</u>	<u>Q</u>	<u>T</u>	<u>G</u>	<u>K</u>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	<u>A</u>	<u>L</u>	<u>K</u>	<u>D</u>	<u>A</u>	<u>V</u>	<u>N</u>	
T. a.	<u>R</u>	<u>N</u>	<u>P</u>	<u>Q</u>	<u>T</u>	<u>K</u>	<u>K</u>	<u>V</u>	<u>I</u>	<u>E</u>	<u>V</u>	<u>P</u>	<u>S</u>	<u>K</u>	<u>K</u>	<u>K</u>	<u>F</u>	<u>V</u>	<u>F</u>	<u>R</u>	<u>A</u>	<u>S</u>	<u>S</u>	<u>K</u>	<u>I</u>	<u>K</u>	<u>Y</u>	<u>Q</u>	<u>Q</u>		
SECONDARY STRUCTURE	----- ARM REGION -----																	*	STRAND 3				*	ALPHA 3				*			

Fig. 3 Amino acid sequence data for DNA-binding protein II using the one-letter code. The abbreviations for each are as follows: B. st., *Bacillus stearothermophilus*; E. c.1, *Escherichia coli* NS1 (HU α); E.c.2, *E. coli* NS2 (HU β); C.p., *Clostridium pasteurianum*; R.m., *Rhizobium meliloti*; P.a., *Pseudomonas aeruginosa*; T.a., *Thermoplasma acidophilum*. All homologies are with respect to the *B. stearothermophilus* sequence. Identical amino acids are underlined and written in large capitals. The distribution of secondary structure has been deduced from our electron density map.

the unfavourable orientation of the N-H and the C=O bonds, remain separate and do not form a six-stranded sheet. The two α 3 helices are disposed on either side of the sheet region and the two arms peel off from this tightly packed body.

The dimer has the hydrophobic interior and hydrophilic surface typical of soluble proteins. The amino acids responsible for the hydrophobic core and therefore the dimer stability are highly conserved. The most important are the four phenylalanine residues (29, 47, 50 and 79), leucine 6, isoleucine 32, leucine 36 and leucine 44 (see Fig. 3). NMR studies have indicated the hydrophobic interactions of the phenylalanines³⁰.

Hydrodynamic studies have suggested that the protein exists in solution either as a tetramer or an elongated dimer^{9,15,20,21}. Without prior knowledge of the unusual shape of the dimer with its two extended arms, it would be impossible to interpret correctly the results from such experiments. We believe that the extensive monomer-monomer contacts, in particular the hydrophobic interactions at the core, clearly point to the dimer as the fundamental unit of the protein. It is also possible that the dimer is in equilibrium with forms of higher aggregation, as suggested

by cross-linking experiments^{9,20,31}, but there are no dimer-dimer contacts in the crystal suggestive of a possible higher-order structure.

Protein-DNA interaction

The visible parts of the arms together with the two symmetry-related strands 3 between them create a concave surface on the dimer which is exactly complementary to the right-handed double helix of B-DNA. This helical depression has a diameter of ~25 Å and this indicates that the protein interacts electrostatically with the DNA via the sugar-phosphate backbone. Independent data support this general model. First, the amino acid sequence in the arm is highly conserved (Fig. 3), which indicates that it has a vital functional role. Second, the arm contains four of the five arginines in the molecule and NMR studies on the oligonucleotide binding have suggested that arginines are involved in the binding mechanism^{15,31}. Third, the protein binds double-stranded DNA nonspecifically and can also bind, albeit less strongly, RNA⁹ and a whole range of single- and double-stranded oligonucleotides¹⁵. Thus, it is the sugar-phosphate

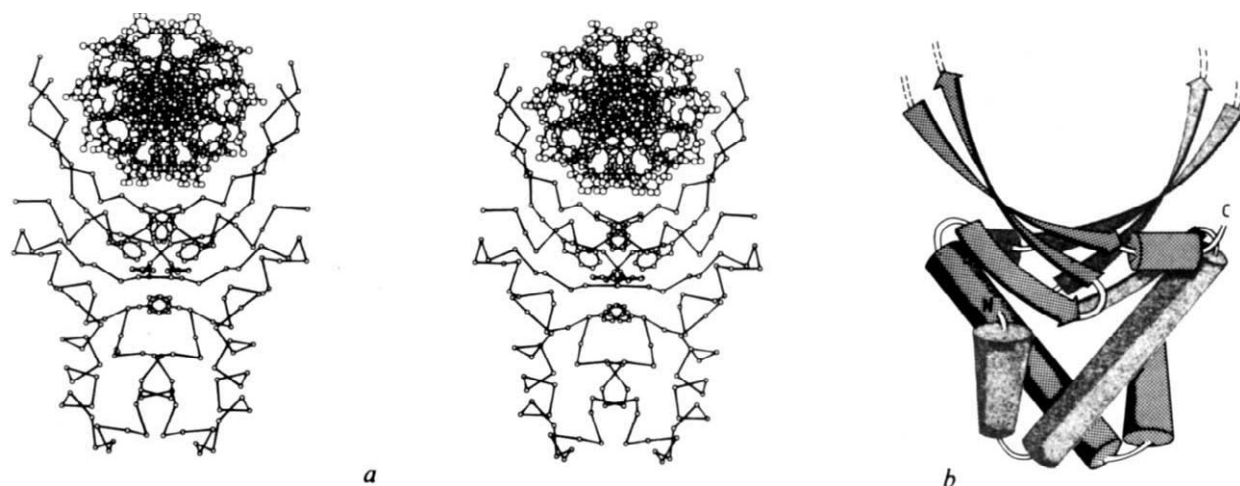


Fig. 4 Structure of the dimer of DNA-binding protein II. *a*, A stereo view of the C α backbone together with its putative interaction with DNA. The distal regions of the arms which are proposed to wrap around the DNA are missing because they are not visible in the electron density map. The phenylalanine side chains which are important for the stabilization of the dimer in the core of the molecule are shown. *b*, A schematic representation of the secondary structure. For clarity, each monomer is shaded differently.

backbone that appears to be important for the interaction. Finally, as would be expected for a simple electrostatic mechanism, all binding is totally inhibited by high salt concentrations⁸.

We have attempted to study the binding directly using various oligonucleotides in crystal soaking and co-crystallization experiments. These have so far been unsuccessful and a possible reason for this is that the major dimer-dimer contact in the crystal involves the helical base of one molecule occupying the proposed DNA-binding region of the molecule below. Therefore, in the former experiments the site would be blocked and in the latter a major interaction responsible for possible crystal formation would not be available. We therefore used model building on the computer graphics to study the proposed DNA-binding mechanism.

Using the protein coordinates and a standard set of coordinates for the B form of the DNA double helix³², we investigated how the rigid molecules might dock together. In common with previous similar studies^{27-29,33}, we recognized the importance of first aligning the protein 2-fold axis with the pseudo-2-fold axis of the DNA perpendicular to the helix axis. On one side of the helix this axis bisects the minor groove and on the opposite side the same axis bisects the major groove. Depending on the side at which the protein has access to the DNA, two distinct models for the interaction are obtained. In the first, the arms follow the path of the minor groove and in the second they follow the major groove. Both possibilities were optimized by rotating and translating the protein relative to the DNA until good van der Waals contacts were made.

Although the position of the DNA-binding arms has been clearly defined, the distinction between the two models is very small, and a definite choice is not yet possible. In both, the long flexible arginines would be able to interact with phosphates of the DNA backbone, and the natural twist of the β -ribbon, together with the β -bulges, would bend the arm in the direction of the DNA helical path. Also, the arms are of the correct length to enable one dimer to cover one turn of the DNA double helix. This is important when we consider how the molecule might induce a supercoiling of the DNA (see below). Figures 5 and 6 show the model with the arms following the major groove, but this is a purely arbitrary choice.

It has been noted³⁴ that an antiparallel β -ribbon has the correct dimensions to fit into the minor groove of B-DNA and interact with the antiparallel sugar-phosphate strands. However, we consider such a mechanism unlikely in this case for two reasons. First, the diameter of the binding site is relatively large and this argues against an intimate contact between the arms and the minor groove. Second, reference to Fig. 3 shows a

definite asymmetry between the two strands of the proposed β -ribbon arm structure. The outgoing strand (residues 52-63) not only contains all the arginines in the arm, but is also more highly conserved than the return strand (residues 64-77). It is therefore likely that the outgoing strand is more important in the interaction with DNA.

DNA supercoiling

It has been reported that purified DNA-binding protein II has the ability to form an average of 14 bead-like objects along relaxed, circular simian virus 40 DNA in the presence of a nicking-closing enzyme⁸. During this process, an equivalent number of (negative) superhelical turns are introduced into the DNA. Therefore we conclude that, like the eukaryotic core particle, the beads represent a winding of the DNA around a protein core. However, although there is evidence that the protein can self-associate in solution to form higher aggregates (see above), no chromatin-like core particles have been detected in the absence of DNA. In contrast, the histone proteins do have this ability¹. Therefore, the DNA-protein complex that we have proposed should exhibit features indicative of a mechanism for self-association into a superhelical structure.

One such feature is very obvious and concerns the general shape of the protein. When the complex is viewed normal to both the DNA axis and the common 2-fold axis (Figs 5, 6), the bound protein is strikingly wedge-shaped and therefore well adapted to form part of a circular assembly. Given our observation that one dimer can cover one turn of the DNA helix, one can visualize the type of mechanism illustrated in Fig. 6. The important point is that several dimers bound adjacently along the helix would be ideally oriented to close into a helical array with the DNA bound on the outside. The helicity would result from the ways in which the dimers make contacts with one another.

As with the DNA-binding mechanism, there is evidence for this supercoiling model in the protein sequence. Apart from those regions which we believe to be important in DNA binding and in the maintenance of the general structure, three other regions are almost totally conserved. These are the N-terminus (residues 1-3), the $\alpha 3$ helix (residues 84-89) and the region including the putative turn at the end of the arm (residues 60-65). All three can be reasonably explained by the model. The dimer-dimer interaction shown in Fig. 6 would bring the $\alpha 3$ helix of one into close contact with the N-terminus of the next. Indeed, there is a pocket in the molecule adjacent to the N-terminus and directly above strands 1 and 2 into which the helix could conceivably fit. An explanation for the third

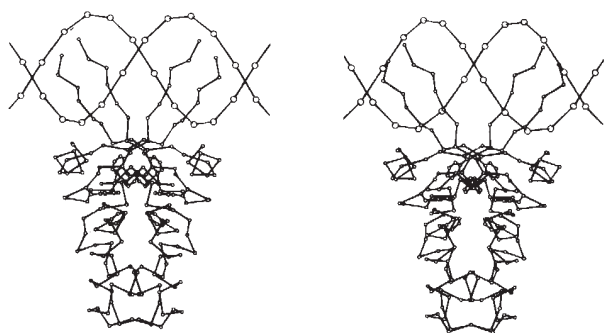


Fig. 5 A stereo view of the proposed interaction between DNA-binding protein II and the double helix of B-DNA. For clarity, only the C α backbones of each molecule are shown together with the phenylalanine residues at the core of the protein. The arms which follow the DNA major groove are not complete because the distal regions are not visible in the electron density map. An alternative interaction in which the arms follow the minor groove is equally valid (see text).

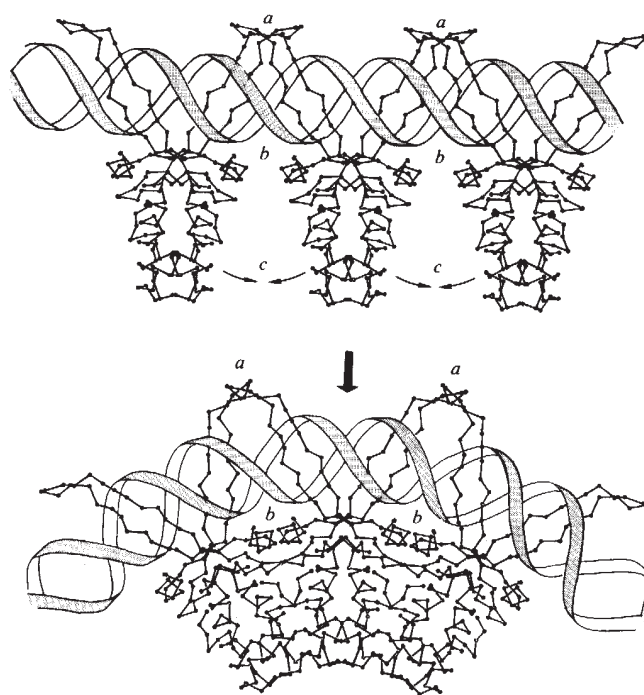


Fig. 6 The proposal for how DNA-binding protein II induces a supercoiling of the DNA. This is merely a diagrammatic representation and does not reflect a detailed structural analysis. The conformations of the distal regions of the arms are included in the diagram. These parts of the protein are not visible in the electron density map and the structures are based on an analysis of the amino acid sequence. The important interactions are labelled. *a*, Contact between the ends of the arms of adjacent molecules; *b*, contact between the C-terminal helix (α -3) and the N-terminus of adjacent molecules; *c*, electrostatic interactions between the wedge-shaped backbones. It is estimated that 8–10 protein dimers and 80–100 base pairs would generate 1 turn of the DNA supercoil. The DNA is not supercoiled here to this extent to avoid a confusing superimposition of the protein dimers.

region is readily apparent from Fig. 6, where it can be seen that the arms from adjacent dimers make intimate contact behind the DNA helix. Such an interaction between the arms would make the complex considerably more stable.

Estimates concerning the size and protein/DNA stoichiometry within the beads vary considerably. Rouvière-Yaniv and co-workers⁸ have measured an average diameter of 180 Å and a protein–DNA content of 8–10 dimers and 275 base pairs. Searcy and Stein³⁵ estimate a diameter of 50–60 Å and suggest a model containing 4 protein molecules and 40 base pairs. Our proposal would generate an object containing approximately 8–10 dimers and therefore 80–100 base pairs per turn, with a diameter of ~140 Å. This is close to the superhelical parameters of DNA in the eukaryotic core particle³. Note that in Fig. 6, for reasons of clarity, the DNA is not supercoiled to this extent and the dimers can make a closer contact than is apparent from the diagram.

Considering the hydrophilic nature of the protein exterior, the probable nature of the dimer–dimer binding is electrostatic and the distribution of the charged residues does appear to divide the surface in question into positive and negative halves to facilitate this. Also, our model of supercoiling requires that the dimer–dimer interaction be weaker than that of the dimer to DNA to allow the latter to complex first. It is therefore interesting that at intermediate salt concentrations the protein can bind to DNA while being unable to induce supercoiling⁸.

Discussion

The structures of three genome regulatory proteins have recently been determined, and they seem to have a common sequence-specific DNA-binding mechanism³⁶: *cro* (ref. 27) and the DNA-binding domains of catabolite gene activator protein (CAP)²⁸ and λ repressor²⁹ each have a pair of symmetry-related helices (α 3 in *cro* and λ repressor and α F in CAP) which are oriented so as to fit into successive major grooves of the B-DNA. Also, the helical units α 2– α 3 in *cro* and λ repressor and α E– α F in CAP are remarkably similar in structure and, to a lesser extent, in their amino acid sequences and this is thought to indicate the central functional role of the unit in this type of protein³⁶.

Clearly, such a DNA-binding mechanism is quite different from that which we have proposed for our structure. However, when one recognizes the fundamental difference between these proteins and DNA-binding protein II, the need for two types of mechanism can be rationalized. The former requires a sequence-specific DNA-binding mechanism and uses a more intimate protein–DNA contact in order to 'read' the sequence of base pairs. The latter requires a non-specific mechanism and therefore uses the invariant sugar–phosphate backbone. There

is evidence that *cro*, CAP, λ repressor and genome regulatory proteins in general use nonspecific DNA binding in a facilitated transfer process which enables them to reach their target sites more quickly³⁷. This binding also seems to involve the sugar–phosphate backbone of the DNA³⁸.

More difficult to rationalize is an apparent structural similarity between the α 1–turn– α 2 region of our protein and the two-helical units of *cro*, CAP and λ repressor which are considered central to their function. A preliminary comparison with *cro*³⁹ revealed that the correspondence is similar to that which exists between *cro* and λ repressor⁴⁰ and *cro* and CAP⁴¹. The particular arrangement of these helical units about the 2-fold axis which makes them such obvious candidates for the site of DNA binding in the regulatory proteins is not present in our structure and a functional equivalence is, therefore, not likely. However, it is interesting that no other known protein structure contains a similar unit³⁶, and we intend to undertake more detailed comparisons of the structures and sequences.

Two other proteins, prealbumin³³ and gene 5 DNA-binding protein⁴², have regions similar to that which we have described. Both contain arm structures and a connecting sheet region which are thought to interact with DNA. However, prealbumin has never been shown to bind DNA and gene 5 protein binds single-stranded DNA.

Functionally, DNA-binding protein II is most closely related to the histone proteins and more specifically to H3 and H4: these are rich in arginine residues and there is evidence that they are the first to bind DNA in the core particle formation⁴³. When the structures of these proteins do become available, it

will be interesting to look for any structural similarities and any signs of an evolutionary relationship which are not immediately apparent from the amino acid sequence.

We thank Professor H.-G. Wittmann for his support of this work, Volker Brachvogel and Howard Terry for technical assistance, Professor Wolfram Saenger and Rolf Hilgenfeld for stimu-

lating discussion and for making their molecular graphics system available, and their students for building a molecular model of the protein. We also thank Richard Brimacombe and Eric Dabbs for reading the manuscript. This work was in part funded by the Deutsche Forschungsgemeinschaft (SFB9). I.T. is a Humboldt Fellow.

Received 19 March; accepted 13 June 1984.

1. Sperling, R. & Wachtel, E. J. *Adv. Protein Chem.* **34**, 1–60 (1981).
2. Igo-Kemenes, T., Hoerz, W. & Zachau, H. G. A. *Rev. Biochem.* **51**, 89–121 (1982).
3. Richmond, T. J., Finch, J. T. & Klug, A. *Cold Spring Harb. Symp. quant. Biol.* **47**, 493–501 (1982).
4. Burlingame, R. W., Love, W. E. & Moudrianakis, E. N. *Science* **223**, 413–414 (1984).
5. Worcel, A. & Burgi, E. J. *molec. Biol.* **71**, 127–147 (1972).
6. Griffith, J. D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 563–567 (1976).
7. Geider, K. & Hoffmann-Berlin, H. A. *Rev. Biochem.* **50**, 233–260 (1981).
8. Rouvière-Yaniv, J. & Yaniv, M. *Cell* **17**, 265–274 (1979).
9. Berthold, V. & Geider, K. *Eur. J. Biochem.* **71**, 443–449 (1976).
10. Rouvière-Yaniv, J. & Gros, F. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3428–3432 (1975).
11. Searcy, D. G. *Biochim. biophys. Acta* **395**, 535–547 (1975).
12. Haselkorn, R. & Rouvière-Yaniv, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1917–1920 (1976).
13. Hawkins, A. R. & Wootton, J. C. *FEBS Lett.* **130**, 275–278 (1981).
14. Imber, R., Baechinger, H. & Bickle, T. A. *Eur. J. Biochem.* **122**, 627–632 (1982).
15. Dijk, J., White, S. W., Wilson, K. S. & Appelt, K. J. *biol. Chem.* **258**, 4003–4006 (1983).
16. Laine, B., Bélaiche, D., Khanaka, H. & Sautière, P. *Eur. J. Biochem.* **131**, 325–331 (1983).
17. Mende, L., Timm, B. & Subramanian, A. R. *FEBS Lett.* **96**, 395–398 (1978).
18. DeLange, R. J., Williams, L. C. & Searcy, D. G. *J. biol. Chem.* **256**, 905–911 (1981).
19. Kimura, M. & Wilson, K. S. *J. biol. Chem.* **258**, 4007–4011 (1983).
20. Suryanarayana, T. & Subramanian, A. R. *Biochim. biophys. Acta* **520**, 342–357 (1978).
21. Rouvière-Yaniv, J. & Kjeldgaard, N. O. *FEBS Lett.* **106**, 297–300 (1979).
22. Vashavsky, A. J., Nedospasov, S. A., Bakayev, V. V., Bakayeva, T. G. & Georgiev, G. P. *Nucleic Acids Res.* **4**, 2725–2745 (1977).
23. Rouvière-Yaniv, J. *Cold Spring Harb. Symp. quant. Biol.* **42**, 439–447 (1978).

24. Thomm, M., Stetter, K. O. & Zillig, W. *Zbl. Bakt. Hyg., I Abt. Orig. C3*, 128–139 (1982).
25. Busby, S., Kolb, A. & Buc, H. *Eur. J. Biochem.* **99**, 105–111 (1979).
26. Tanaka, I., White, S. W., Appelt, K., Wilson, K. S. & Dijk, J. *FEBS Lett.* **165**, 39–42 (1984).
27. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 754–758 (1981).
28. Steitz, T. A., Weber, I. T., Ollis, D. & Brick, P. *J. biomolec. Struct. Dyn.* **1**, 1023–1037 (1983).
29. Pabo, C. O. & Lewis, M. *Nature* **298**, 443–447 (1982).
30. Paci, M., Pon, C. L., Losso, M. A. & Gualerzi, C. O. *Eur. J. Biochem.* **138**, 193–200 (1984).
31. Lammi, M., Paci, M. & Gualerzi, C. O. *FEBS Lett.* **170**, 99–104 (1984).
32. Arnott, S. & Hukins, D. W. L. *Biochem. biophys. Res. Commun.* **47**, 1504–1510 (1972).
33. Blake, C. C. F. & Oatley, S. J. *Nature* **268**, 115–120 (1977).
34. Church, G. M., Sussman, J. L. & Kim, S.-H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1458–1462 (1977).
35. Searcy, D. G. & Stein, D. B. *Biochim. biophys. Acta* **609**, 180–195 (1980).
36. Takeda, Y., Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. *Science* **221**, 1020–1026 (1983).
37. Berg, O. G., Winter, R. B. & von Hippel, P. H. *Trends biochem. Sci.* **7**, 52–55 (1982).
38. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., Fisher, R. G. & Takeda, Y. *Trends biochem. Sci.* **8**, 25–29 (1983).
39. Ohlendorf, D. H., Anderson, W. F., Takeda, Y. & Matthews, B. W. *J. biomolec. Struct. Dyn.* **1**, 553–563 (1983).
40. Ohlendorf, D. H., Anderson, W. F., Lewis, M., Pabo, C. O. & Matthews, B. W. *J. molec. Biol.* **169**, 757–769 (1983).
41. Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. & Matthews, B. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3097–3100 (1982).
42. Brayer, G. D. & McPherson, A. *J. molec. Biol.* **169**, 565–596 (1983).
43. Camerini-Otero, R. D., Sollner-Webb, B. & Felsenfeld, G. *Cell* **8**, 333–347 (1976).

Novel stereospecificity of the L-arabinose-binding protein

Florante A. Quijcho & Nand K. Vyas

Department of Biochemistry, Rice University, Houston, Texas 77251, USA

Tertiary structure refinement at 1.7 Å resolution of the liganded form of L-arabinose-binding protein from *Escherichia coli* has revealed a novel binding site geometry which accommodates both α - and β -anomers of L-arabinose. This detailed structure analysis provides new understanding of protein–sugar interaction, the process by which the binding protein minimizes the difference in the stability of the two bound sugar anomers, and the roles of periplasmic binding proteins in active transport.

THE modes of binding of ligands have a profound and specific influence over the structure and function of proteins. In order to establish the details and obtain definitive conclusions concerning protein–ligand interactions it is essential that an accurate and unbiased description of the molecular geometry is determined at high resolution by crystallographic refinement methods. We have recently used the restrained least-squares method of Konnert and Hendrickson¹ to refine extensively the structure of the liganded form of the L-arabinose-binding protein (ABP) at 1.7 Å resolution. Here we focus on the most important finding of our study: the unique molecular architecture of the binding site which provides identical binding properties for both α and β anomers of L-arabinose. Moreover, several features of the complex are relevant to a general understanding of ligand–protein interaction.

L-Arabinose-binding protein is a member of a family of proteins found in the periplasm of Gram-negative bacteria. These proteins are essential components of osmotic shock-sensitive active transport systems for a large variety of carbohydrates, amino acids and ions^{2,3}. Furthermore, several of the sugar-binding proteins serve as receptors in bacterial chemotaxis⁴. Protein components residing in the cytoplasmic membrane are also required for both cellular processes^{3,4}.

Like all periplasmic binding proteins, ABP consists of a single polypeptide chain with one high-affinity binding site. In addition to ABP from *Escherichia coli*⁵, we have recently determined the molecular structures of the sulphate-binding protein from *Salmonella typhimurium*⁶, and leucine, isoleucine, valine-binding protein⁷ and D-galactose-binding protein⁸ from *E. coli*. Despite the lack of significant sequence homology, these proteins have a high degree of structural similarity. They are elongated and composed of two globular domains connected by three separate polypeptide segments; each domain contains a central β -pleated sheet sandwiched between helices. These structural features are also found in the *S. typhimurium* D-galactose-binding protein⁹. Moreover, all these proteins have a common ligand binding site located in the cleft between the two domains^{6–10}.

Refinement and assessment of structure

Purified ABP contains one molecule of tightly bound L-arabinose^{10,11}. The amino acid sequence indicates a molecular weight of 33,170 for 306 residues¹². Single crystals of ABP, which diffract to better than 1.7 Å resolution, belong to the space group P2₁2₁2₁ with unit cell dimensions $a = 55.46$ Å, $b = 71.82$ Å, $c = 77.84$ Å.

Fig. 1 a, Superimposed difference electron density maps of the α - and β -L-arabinose bound to ABP. The density and refined model of the α -anomer are shown in blue and the β -anomer is shown in purple. Coincidence of both sugar densities (or both colours) appears as white. The α -OH is labelled O(1)A and the β -OH labelled O(1)B. The difference map of one anomer is independently computed from a conventional $(|F_o| - |F_c|, \alpha_c)$ residual map in which the calculated contribution of the same anomer at 50% occupancy is omitted from F_c . Superimposed maps of both anomers were then displayed on the graphic system. The electron density surface (grid spacing 0.25 Å) is 0.16 electrons Å⁻³, approximately three times the standard deviation of the final difference map. Note that the density of all the hydroxyls in both anomers are comparable, an indication that each anomer refined to about half occupancy. The separation of the two sugar positions is better shown in **b**. **b**, The same as **a** except that the sugar density and model have been oriented differently and 'clipped' in the 'front' in order to show more clearly the separation between the two sugar positions and the accurate fit of the refined sugar model to the corresponding density. The term 'front' refers to the stereo direction, towards the reader.

Fig. 2 a, The refined molecular model (purple) of the ligand site of ABP with bound α - or β -L-arabinose superimposed on the electron density contour surfaces (blue). The map was calculated with coefficient $2|F_o| - |F_c|$ and calculated phases. The density surfaces (grid-spacing 0.40 Å) were contoured at 1 electron Å⁻³, 1.16 times the estimated standard deviation of the electron density map. In this map, both anomeric hydroxyl densities have equal peak heights but only half of the other unresolved sugar hydroxyls. The hydrogen bonds (indicated by green dashed lines) to the sugars are: Asp 90 to either O(1) anomeric hydroxyl, Lys 10 to O(2), Glu 14 and Asn 205 to O(3), Asn 232 to O(3) and O(4), Arg 151 to O(4) and O(5) ring oxygen, Wat 309 to O(2), and Wat 310 to O(5). See text for further descriptions of the hydrogen-bond networks. **b**, Identical to **a**, but with electron density surface removed to show more clearly the hydrogen-bonding system formed between the sugars and side-chain residues. Including the three hydrogen bonds, O(3) is tetrahedrally coordinated. Asn 232 OD1 and ND2, O(3) and O(4) form a plane with an average deviation of 0.04 Å. Likewise, NH2 and NH1 of Arg 151, O(4) and the sugar ring oxygen O(5) form a plane (average deviation 0.02 Å). The six amino acid residues hydrogen-bonded to the sugars are equally divided between the two domains; the N-terminal domain supplies Lys 10, Glu 14 and Asp 90, whereas the C-terminal domain provides Arg 151, Asn 205 and Asn 232.

With starting atomic coordinates obtained from the 2.4-Å multiple isomorphous replacement (MIR) map⁵, ABP was refined in stages of resolution—1.9, 1.85, 1.75 and finally 1.7 Å. Discrepancies between electron density maps and superposed refined coordinates were manually minimized using FRODO¹³, extensively modified for an Evans and Sutherland PS 300 graphics system¹⁴. Table 1 summarizes specific conditions for refining the bound sugars. The final *R*-factor was 0.137 for 26,634 reflections, with intensities exceeding 3 σ from 8.0 to 1.7 Å resolution.

Pertinent refinement statistics are shown in Table 1. Because the binding site region has one of the lowest average isotropic *B*-factor and r.m.s. coordinate shifts, the accuracy of the coordinates of this region is considerably better than the overall estimate indicated in Table 1. A more detailed presentation of the refinement and a complete description of the entire refined protein structure, including solvent molecules, will be presented elsewhere (N.K.V. and F.A.Q., manuscript in preparation).

Stereospecificity of the binding site

An L-arabinopyranose model, in the *C1* full-chair conformation, was initially fitted to the electron density of the bound sugar in

the 2.4-Å resolution map calculated from MIR phases¹⁰. At this resolution only the β -anomeric form of the sugar was fitted and the presence of the α -anomer could not be established. However, stopped-flow rapid kinetic measurements in solution indicated that ABP, as well as two other sugar-binding proteins, binds both sugar anomers reversibly with the same affinity and at equal rates^{15,16}. Refinement of the structure has now conclusively shown that either α - or β -L-arabinose can be bound in the same site. This analysis has further revealed a novel protein site geometry for binding both sugar anomers.

Superimposed difference Fourier maps (Fig. 1) demonstrate that the density of individual L-arabinose anomers is well defined and clearly differentiated, and that both sugar models fitted their respective density extremely well. The structures of both bound sugars refined to the standard *C1* full-chair conformation commonly observed in crystal structures of simple monopyranosides. This work represents the first determination of the structure of α -L-arabinose.

The atoms common to both anomers are not all identically positioned within the binding site. The positions of C-4 and adjacent atoms are superimposable for the two sugar molecules but, from this hinge, the two sugar positions gradually diverge

Fig. 3 Proximity of Trp 16 and Phe 17 of ABP to a hydrophobic patch in the bound L-arabinose. This stereo pair shows the molecular surface as dots and the stick model of the refined two sugars and side chains. The individual van der Waals' radii used were: C = 1.7 Å, N = 1.6 Å, O = 1.5 Å. Coloured dots correspond to the following atoms: C, green; N, blue; O, red.

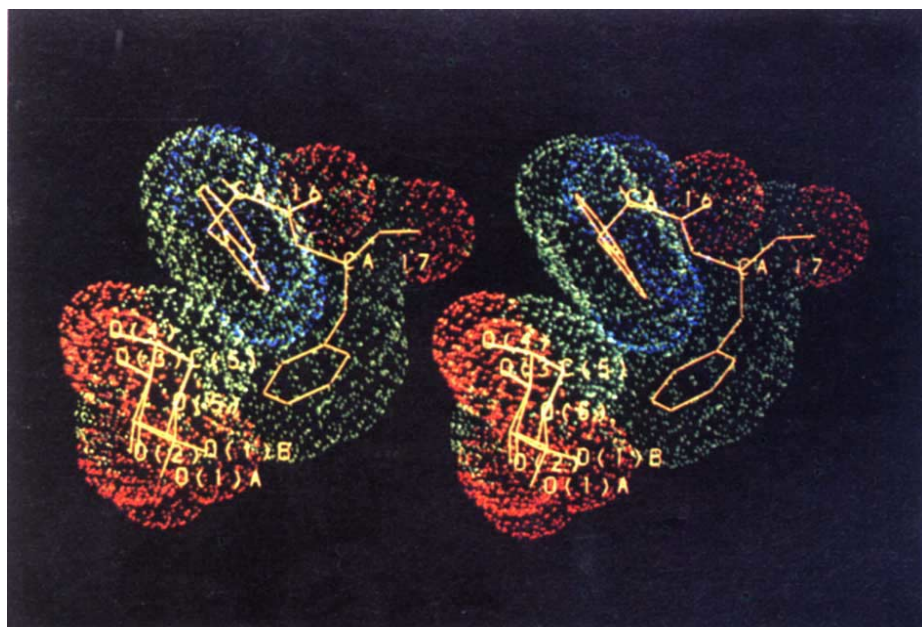


Table 1 Summary of the 1.7 Å restrained least-squares refinement and other data of L-arabinose-binding protein

	Target σ	Final value
<i>R</i> -factor		0.137
r.m.s. coordinate shift (Å)		0.017
r.m.s. <i>B</i> shift (Å ²)		0.13
Average ($ F_o - F_c $)	(18, -45)	25.53
Bond distance (Å)	0.015	0.016
Angle distance (Å)	0.030	0.044
Planarity (Å)	0.010	0.009
Chiral volume (Å ³)	0.05	0.07
Angle ω (deg.)	5.0	3.2
r.m.s. coordinate error (Å)*		0.14
No. of reflections at 8–1.7 Å		26,634
No. of atoms refined:		
Protein		2,316
Two molecules of L-arabinose†		20
Water molecules		230

Where listed together, the values of target σ are the inverse square root of the least-squares weight and corresponding final values given are the r.m.s. deviations of the parameters from their respective ideal values taken from the standard groups listed in Table 2 of ref. 32.

* Estimated by Luzzati plot³³.

† The initial fit of β -L-arabinose¹⁰ was first included in the refinement at 1.85 Å resolution. The bond lengths, angles and torsions of the sugar were restrained to values derived from the known β -DL-arabinose structure³⁴. The presence of the α -L-arabinose in the same site was later unambiguously established from Fourier and difference Fourier maps calculated following refinement at 1.75 Å (*R*-factor, 0.18). The initial coordinates of α -anomer atoms were identical to those of β -anomer with the exception of O-1 which was obtained from the corresponding peak in the difference Fourier map. In subsequent refinement of both anomers, starting with both sugar coincident, the following conditions were implemented. (1) The ideal values for the α -anomer stereochemical restraint parameters were identical to those for the β -anomer, except that O-1 was derived after rotation of the axial anomeric hydroxyl to the equatorial position. (2) There were no repulsive restraints for non-bonded contacts, including hydrogen-bond distances, between the two sugar molecules and between the sugar and the protein. (3) Restraints on chiral volumes of asymmetrical carbon atoms were additionally imposed on both anomers. (4) Atoms for both anomers were given an initial occupancy of 0.5 and the initial thermal parameters for individual atoms of the α -anomer were those obtained from the previously refined β -anomer; each parameter was refined in alternate cycles. Because occupancies of the atoms of each anomer deviated little, they are eventually fixed at the initial values of 0.5. This is consistent with the almost equal electron density of both anomeric hydroxyls in maps shown in Figs 1 and 2. (5) In the final stage of refinement, only one sugar was refined until the r.m.s. shift in coordinates had reached a minimum value. The other sugar was then refined. This alternate refinement was repeated until the r.m.s. coordinate shift for both sugars finally converged at 0.007 Å, significantly lower than the value for the entire structure, indicated above.

roughly in the direction parallel to the *b* axis (see Fig. 1*b*). This divergence produces a maximal separation of 0.31 Å in the positions of the C-1 atoms and brings the anomeric hydroxyls to within 1.84 Å of each other. If both pyranose rings had identical orientations, the difference in positions of the anomeric hydroxyls would be 2.3 Å.

Sugar-protein interaction

Details of the interaction between the anomeric sugars and protein residues are presented in Fig. 2 and Table 2 and summarized in Fig. 4. The result shown in Fig. 2 leaves little doubt as to the reliability and accuracy of the analysis of the sugar-ABP interaction. This interaction is far more complex and extensive than initially deduced at 2.4 Å resolution¹⁰.

Although both anomers of L-arabinose bind, the geometry of the ligand site remains the same and identical amino acid side chains interact with both sugars. An important discovery is that Asp 90 OD2 is precisely aligned to accept a hydrogen bond from either the α - or β -anomeric hydroxyl. Lys 10 NZ makes

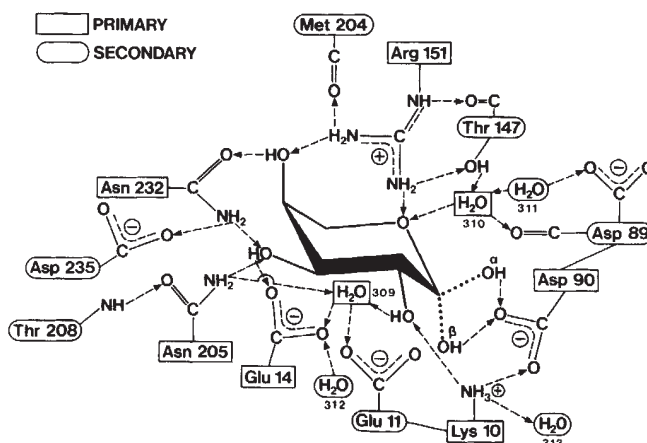


Fig. 4 Schematic diagram of the complex networks of hydrogen bonds formed between L-arabinose and side-chain residues in the binding site region of ABP. This diagram is not meant to indicate that both pyranose rings coincide (see Fig. 1). The residues (primary) that interact with the sugar are in turn hydrogen-bonded to each other or to other (secondary) residues or isolated water molecules. Wat 309 links Glu 14 of the N-terminal domain with Asn 205 of the C-terminal domain. Wat 310 is part of a chain of interactions involving Asp 69–Asp 90 of the N-terminal domain and Thr 147 and Arg 151 of the C-terminal domain.

a hydrogen bond with O-2 and stabilizes the alignment of Asp 90 through a salt link. Lys 10 is also within van der Waals' distance (3.2 Å) of either O-1. As α -OH and β -OH are related by a local mirror plane formed by Lys 10 NZ, Asp 90 OD1, CG and OD2, and a point midway between C-1 of both sugars (mean deviation from the plane, 0.05 Å), each anomeric hydroxyl is involved in equivalent interactions with atoms in the plane (see Fig. 2*b*). Glu 14 OE2 is a recipient of a hydrogen bond from O-3 which in turn is an acceptor from Asn 205 ND2. Unusual bidentate hydrogen bonds are formed between ND2 and OD1 of Asn 232 and O-3 and O-4, respectively, and between NH2 and NH1 of Arg 151 and O-4 and O-5, respectively (see Fig. 2*b*). Two isolated water molecules (Wat 309 and Wat 310) form hydrogen bonds with O-2 and O-5 and bring the sugar and the two domains together in the cleft (see Fig. 4).

Several additional properties of the hydrogen bonds between sugar and protein are worth noting (Figs 2, 4). (1) The anomeric hydroxyls participate solely as hydrogen-bond donors whereas the rest of the sugar hydroxyls serve as both donors and acceptors. This pattern is also observed in the crystal packing of carbohydrates¹⁷. (2) The hydrogen-bonding capacity of O-3 is fully utilized by donating one and accepting two hydrogen bonds. As depicted in Fig. 2*b*, the orientations of the three hydrogen bonds make O-3 approximately tetrahedrally coordinated, including the C–O bond (average angle, $108 \pm 13^\circ$). (3) Based on the geometry (see Fig. 2*b*), the hydrogen bond donated to O-2 or O-4 is shared by both lone pairs of electrons on either sugar hydroxyl. (4) The ring-oxygen is almost tetrahedrally coordinated, hence each of the two ring-oxygen *p* orbitals is directed towards the respective hydrogen-bond donor. (5) All the potential hydrogen-bond donors from the sugars and side chains of Lys 10, Asn 205, Asn 232, Arg 151 and the two water molecules (Wat 309 and Wat 310) are fully utilized.

Because the sugars provide cohesive interactions between the two domains in the cleft region, they stabilize the closed conformation proposed for the liganded form of the protein^{10,18}. The six (primary) amino acid residues hydrogen-bonded to the sugars are equally divided between the opposite walls of the cleft (Fig. 2*b*). Furthermore, as shown schematically in Fig. 4, these (primary) residues are in turn involved in networks of hydrogen bonds with other (secondary) side chains and isolated

Table 2 Hydrogen bonds associated with the binding of α - and β -L-arabinose to ABP

Hydrogen bonds		α -L-arabinose		β -L-arabinose	
Donor (X)	Acceptor (Y)	X...Y (Å)	X—H...Y (deg)	X...Y (Å)	X—H...Y (deg)
O-1	Asp 90 OD2	2.77	159	2.74	166
O-2	Wat 309 O	2.61	155	2.60	156
O-3	Glu 14 OE2	2.77	176	2.66	173
O-4	Asn 232 OD1	2.62	178	2.62	176
Lys 10 NZ	O-2	2.73	146	2.86	151
Asn 205 ND2	O-3	3.03	169	3.09	167
Asn 232 ND2	O-3	2.97	162	2.98	159
Arg 151 NH2	O-4	2.82	167	2.81	168
Arg 151 NH1	O-5	3.05	161	2.99	160
Wat 310 O	O-5	2.80	—	2.74	—
Overall mean		2.82 (0.15)	164 (10)	2.81 (0.16)	164 (8)

Although hydrogens were not refined, sufficient accurate structural information was available from the 1.7 Å refined structure to deduce the location of hydrogen atoms on the basis of geometrical consideration and to measure distances and angles between hydrogen donor and acceptor atoms. Because ABP crystals were kept at pH 6.4, the carboxyl groups of Glu 14 and Asp 90 were assumed to be fully ionized. Determination of hydrogen atom positions in Arg 151, Asn 205 and Asn 232 were strictly based on standard geometry. On the other hand, the three NZ hydrogens of Lys 10 have been placed assuming a perfect staggering in relation to the neighbouring tetrahedral carbon atoms. A similar approach was used for each of the sugar hydroxyls, except that the initial choice of which of the three staggered positions was occupied by the single hydroxyl hydrogen was made on the basis of the nearest hydrogen-bond acceptor. Moreover, as neutron diffraction studies of simple sugars indicate that hydroxyl hydrogens vary from staggered to eclipsed³⁵, the initial choice was rotated, if necessary, about the H—C—O—H torsional angle in order to attain the most linear hydrogen bond with the acceptor. The hydrogen bonds listed above are within the normally accepted range^{17,35–37}.

solvent molecules. Not shown is a further, third shell of residues that forms hydrogen bonds with the secondary residues. This interlocking hydrogen-bond system not only fixes the orientations of the primary residues confined to their respective domains, but also stabilizes the sugar-protein complex. Thus, the geometry of the protein is unchanged on binding either anomer and the formation of identical interactions to both forms of the sugar is achieved by the slight shift in the positions of the sugars.

The binding site region is one of the least mobile regions of the protein, undoubtedly due to the presence of the sugar substrate. The average refined isotropic temperature factor for all atoms in ABP is 18.2 Å², whereas the average value for the bound sugars, the six amino acid residues and two water molecules hydrogen-bonded to the sugar is 10.3 Å².

Hydrophobic patch

In the L-arabinose molecule, the disposition of both the equatorial OH-3 and axial OH-4 to one side of the ring creates a hydrophobic patch composed of C-3, C-4 and C-5. This patch is adjacent to Trp 16 and Phe 17 (Fig. 3). This unique arrangement corroborates the observation that one of the five tryptophans in ABP accounts for most of the fluorescence spectral change elicited by sugar binding^{15,19}. Moreover, the blueshift in the spectrum, which is suggestive of a more hydrophobic environment, is consistent with the X-ray analysis. Similar fluorescence changes have been observed in other sugar-binding proteins and these have facilitated kinetic studies of ligand binding by the rapid-stopped flow technique^{15,16}.

Accessibility of binding site

A calculation of the surface accessibility²⁰ of the two bound sugars in the refined structure, after removal of all the water molecules, has confirmed the initial finding¹⁰ that both sugar molecules are (completely) sequestered in the cleft between the two globular domains. Both bound sugars have an accessible surface area of 2.1 Å², 2.3% of the values for the free α - or β -anomer in the observed conformations.

Additionally, the six amino acid residues hydrogen-bonded to the sugar are essentially buried even in the absence of both sugar molecules. Although the net charge of these residues is neutral, only Lys 10 and Asp 90 form a salt link together, whereas Glu 14 and Arg 151 are each isolated on either side of the sugar ring and not associated in salt linkages (Figs 2, 4). The negative charge on Glu 14 is substantially neutralized by a helix dipole (composed of residues 14–32) pointing at the side chain (with

opposite polarity) and further dissipated via a channel containing a string of three water molecules that link the residue to the bulk solvent (N.K.V. and F.A.Q., unpublished data). The peptide carbonyl dipoles of Met 204 and Thr 147 are aligned at Arg 151 NH2 and NE, respectively (Fig. 4), but these only partially offset the buried positive charge. As discussed below, the positive charge of Arg 151 is probably an important feature of the sugar-protein complex.

Discussion

Efficient utilization of the open-chain, aldehyde form of L-arabinose for the biosynthesis of pentose phosphates in the microorganism²¹ requires translocation of both sugar anomers by the binding protein-dependent active transport system. This study has provided a complete description of how ABP confers sugar specificity on the system. Moreover, the geometry of the ABP-binding site, despite its complexity, is designed to accommodate precisely either L-arabinose anomer while maintaining the same orientations of the residues in the site. One of the most remarkable discoveries of this analysis is the exact complementarity between each anomer and the binding site region, despite the involvement of several residues from both domains in forming the cleft and the requirement for a conformational change to juxtapose these residues and modulate access to/from the binding site cleft.

Results of sugar-binding studies of ABP in solution can now be rationalized on the basis of the structure of the ligand site. Most notable of these are: (1) Both anomeric sugars bind at equal rates and with tight affinity¹⁵. Expulsion of water molecules, the formation of an extensive hydrogen-bonding system and the occlusion of ligands in the binding site all contribute to the extremely tight ligand binding. Note that the extremely small dissociation constants (K_d) of purified binding proteins *in vitro* (~1 μ M) are comparable with the K_m values for transport *in vivo* of highly polar solutes². Moreover, the rates of ligand dissociation from binding proteins measured *in vitro* correlate with the V_{max} for *in vitro* transport^{15,16}. (2) All free sugar hydroxyls are required for binding, and epimers of L-arabinose and D-galactose do not bind to ABP^{11,22}. (3) Sugar induces a change in tryptophan fluorescence (refs 15, 19 and see above). (4) A residue(s) of $pK \sim 4.5$ is present in the binding site and the pH range of maximal binding activity is broad (pH 5–9)¹⁹.

There seems to be an effect, first described by Edwards²³ and later named the anomeric effect by Lemieux²⁴, which correlates the relative stability of the arrangements at the anomeric carbon

with the separation of the anomeric substituent dipole and the dipole formed by the resultant unshared electron pairs on the ring-oxygen. Thus, because these dipoles form a small angle when the substituent is equatorial and a large one when it is axial, the α -L-arabinose is less stable than the β -anomer. Moreover, because the anomeric effect varies inversely with the dielectric constant, the binding site of ABP would make the α -anomer even more unstable, especially as the α -OH is in close proximity to Thr 147 CG2. ABP apparently neutralizes the anomeric effect, because about equal amounts of both anomers are bound (Table 1, Figs 1, 2). This is accomplished by the interaction of the ring-oxygen with the highly effective positive charge density on the guanidinium group of Arg 151, which more than counteracts or opposes the dipole moment on the ring-oxygen. The anomeric effect is further offset by the highly dipolar water molecule (Wat 310) hydrogen-bonded to the ring-oxygen (Figs 2, 4).

Many features of the sugar-ABP complex will undoubtedly be found in other binding proteins, particularly among the sugar binders that collectively possess very similar properties. For example, the D-galactose- and D-maltose-binding proteins also bind both anomeric sugars with identical rates and tight affinity^{15,16}. Moreover, the locations of the binding site in these two carbohydrate protein receptors (refs 8, 9 and J.C.S. and F.A.Q., unpublished data), and in two others with specificity for sulphate⁶ and leucine, isoleucine or valine⁷, are very similar to the L-arabinose-binding protein. As the D-galactose and D-maltose protein receptors are also components for chemotaxis, their ability to recognize both sugar anomers can generate a rapid chemotactic response.

Müller-Hill's proposal²⁵ that each of the two domains in mature sugar-binding proteins is formed from one continuous polypeptide segment and involved in one particular function (the N-terminal domain for membrane interaction and the C-terminal domain for ligand binding) cannot be justified or supported by structural results. Both domains of arabinose- and galactose-binding proteins consist of two separate polypeptide segments^{5,8,9} and participate in sugar binding⁸⁻¹⁰ (see also Fig. 2). Moreover, it is more likely that both domains are involved in the interaction with membrane protein components^{7,10,18,26}.

The present study is the first direct demonstration of a protein-binding site geometry that can accommodate either anomeric form of a sugar substrate and simultaneously minimize the anomeric effect. We are aware that several enzymes involved in glucose metabolism (for example, hexokinase, glucokinase, glucose 6-phosphatase) also lack anomeric specificity. The possible similarity of the ABP-binding site with the *lac* repressor has also afforded insight into the mode of binding of sugars to the repressor (see accompanying paper²⁷).

In addition to conferring substrate specificity, the major role of periplasmic binding proteins in transport is linked to the extremely tight ligand binding. Transport is facilitated by having at least two juxtaposed protein components in the system which differ sufficiently in ligand affinity, being high in the binding protein on the periplasmic (uptake) site and low in the protein(s) confined to the plasma membrane. Active transport is achieved by translocation of solutes via these components in synchrony with conformational changes that propagate through the entire system. The coupling of energy to the binding protein-dependent transport systems requires a metabolite related to ATP²⁸.

Further importance of binding proteins in active transport is indicated by the results of this and other structural analyses. For instance, ligand-induced cleft closure^{10,18} generates the appropriate stereochemistry for the specific interaction of the liganded protein, in preference to the unliganded form, with the membrane-bound protein components, thus initiating the translocation process. Sequestering of ligands should facilitate transport by excluding the bulk water at the interface between binding proteins and respective membrane components. A bilobate structure, with binding site in the cleft between the two lobes, is an important common feature of binding proteins; it facilitates a hinge-bending motion of the two domains to open and close the cleft^{7,18,26}. This conformational change could affect the interaction with other membrane-bound components, the affinity for substrate, or both. A plausible mode of interaction between the membrane protein aggregate and a bilobate binding protein has been proposed^{7,26}. The membrane proteins, with binding sites for closely related substances^{3,29}, presumably serve as carriers for the translocation of highly polar solutes (ions, carbohydrates, amino acids) across the membrane.

Refinement of the ABP structure has for the first time provided detailed features of the binding of a carbohydrate substrate, with broad implications for our understanding of other similar complexes. Moreover, this analysis, together with other data for ABP derived from a variety of techniques (including NMR³⁰, kinetic^{15,16}, calorimetric³¹) provides an excellent basis for the analysis of ABP with altered activity or specificity generated by site-specific mutagenesis, the design of other substrates and the theoretical study of ligand binding. Finally, this report not only further validates the restrained refinement technique but clearly underscores the importance and, indeed, the necessity of high resolution refinement of protein structures.

We thank Drs J. W. Pflugrath and M. A. Saper for invaluable assistance. This work was supported by grants from NIH (GM21371), NSF (PCM82-14057) and the Robert A. Welch Foundation (C-581). F.A.Q. was a Guggenheim Fellow during part of this work.

Received 9 January; accepted 11 May 1984.

- Hendrickson, W. A. & Konnert, J. H. in *Computing in Crystallography* (eds Diamond, R., Ramaseshan, S. & Venkatesan, K.) 13.01-13.23 (Indian Academy of Sciences, International Union of Crystallography, Bangalore, 1980).
- Wilson, D. B. & Smith, J. B. in *Bacterial Transport* (ed. Rosen, B. P.) 495-557 (Dekker, New York, 1978).
- Landick, R. & Oxender, D. L. in *Membranes and Transport* (ed. Martonosi, A. N.) 81-88 (Plenum, New York, 1982).
- Adler, J. A. *Rev. Biochem.* **44**, 341-356 (1975).
- Gilliland, G. L. & Quijcho, F. A. *J. molec. Biol.* **146**, 341-362 (1981).
- Pflugrath, J. W. thesis, William Marsh Rice Univ. (1983).
- Saper, M. A. & Quijcho, F. A. *J. biol. Chem.* **258**, 11057-11062 (1983).
- Vyas, N. K., Vyas, M. N. & Quijcho, F. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1792-1796 (1983).
- Mowbray, S. L. & Petsko, G. A. *J. biol. Chem.* **258**, 7991-7997 (1983).
- Newcomer, M. E., Gilliland, G. L. & Quijcho, F. A. *J. biol. Chem.* **256**, 13213-13217 (1981).
- Newcomer, M. E., Miller, D. M. III & Quijcho, F. A. *J. biol. Chem.* **254**, 7529-7533 (1979).
- Hogg, R. A. & Hermodson, M. A. *J. biol. Chem.* **252**, 5135-5141 (1977).
- Jones, T. A. in *Computational Crystallography* (ed. Sayre, D.) 3303-3317 (Clarendon, Oxford, 1982).
- Pflugrath, J. W., Saper, M. A. & Quijcho, F. A. in *Computational Crystallography* (ed. Hall, S.) (Oxford University Press, in the press).
- Miller, D. M. III, Olson, J. S., Pflugrath, J. W. & Quijcho, F. A. *J. biol. Chem.* **258**, 13665-13672 (1983).
- Miller, D. M. III, Olson, J. S. & Quijcho, F. A. *J. biol. Chem.* **254**, 2465-2471 (1980).
- Jeffrey, G. A. & Lewis, L. *Carbohydrate Res.* **60**, 179-182 (1978).
- Newcomer, M. E., Lewis, B. A. & Quijcho, F. A. *J. biol. Chem.* **256**, 13218-13222 (1981).
- Parsons, R. G. & Hogg, R. W. *J. biol. Chem.* **249**, 3602-3607 (1974).
- Lee, B. K. & Richards, F. M. *J. molec. Biol.* **55**, 517-526 (1971).
- Engelsberg, E. in *Metabolic Pathways: Metabolic Regulation* Vol. 1 (ed. Vogel, H. J.) 257 (Academic, New York, 1971).
- Hogg, R. W. & Engelsberg, J. *Bact.* **100**, 423-432 (1969).
- Edward, J. T. *Chem. Ind.*, 1102-1104 (1955).
- Lemieux, R. U. in *Molecular Rearrangements* (ed. deMayo, P.) 713-769 (Wiley-Interscience, New York, 1963).
- Müller-Hill, B. *Nature* **302**, 163-164 (1983).
- Quijcho, F. A., Gilliland, G. L. & Phillips, G. N. Jr *J. biol. Chem.* **252**, 5142-5149 (1977).
- Sams, C. F., Vyas, N. K., Quijcho, F. A. & Matthews, K. S. *Nature* **310**, 429-430 (1984).
- Berger, E. A. & Heppel, L. A. *J. biol. Chem.* **249**, 7747-7755 (1974).
- Higgins, C. F. et al. *Nature* **298**, 723-727 (1982).
- Clark, A. F., Gerken, T. A. & Hogg, R. W. *Biochemistry* **21**, 2227-2233 (1982).
- Fukada, H., Sturtevant, J. M. & Quijcho, F. A. *J. biol. Chem.* **258**, 13193-13198 (1983).
- Sielecki, A. R. et al. *J. molec. Biol.* **134**, 781-804 (1979).
- Luzzati, V. *Acta crystallogr.* **5**, 802-810 (1952).
- Kim, S. H. & Jeffrey, G. A. *Acta crystallogr.* **22**, 537-545 (1967).
- Takagi, S. & Jeffrey, G. A. *Acta crystallogr.* **B33**, 3033-3040 (1977).
- Hagler, A. T., Huler, E. & Lifson, S. *J. Am. Chem. Soc.* **96**, 5319-5325 (1974).
- Donohue, J. in *Structural Chemistry and Molecular Biology* (eds Rich, A. & Davidson, N.) 443-465 (Freeman, San Francisco, 1968).

Genomic organization and sequence of T-cell receptor β -chain constant- and joining-region genes

Nicholas R. J. Gascoigne, Yueh-hsiu Chien, Daniel M. Becker,
Joshua Kavalier & Mark M. Davis

Department of Medical Microbiology, School of Medicine, Stanford University, Stanford, California 94305, USA

The genomic structure of the joining (J) and constant (C) regions of the locus encoding the β -chain of the murine T-cell receptor has been analysed. The gene segments are arranged tandemly (J-C/J-C) within a 15-kilobase region. The two constant-region genes are almost identical, differing by only four amino acids, all in carboxy-terminal portions. Each C region comprises four exons encoding an external globular domain, a small hinge-like region, a transmembrane region and a cytoplasmic tail plus 3'-untranslated region. The two clusters of J regions each contain 7 distinct elements, 12 of which may be functional.

T CELLS are both effectors and regulators of the immune system. Helper T (T_H) cells augment the responses of cytotoxic T(T_C) cells and B cells, while suppressor T(T_S) cells reduce these responses. The T_C cells are responsible for the destruction of cells infected by intracellular parasites and also of tumour cells. Like B cells, T lymphocytes have a surface receptor that allows them to distinguish between self and non-self. Unlike the B-cell receptor (antibody), this receptor can only recognize antigen in the context of molecules of the major histocompatibility complex (MHC; HLA in man, H-2 in mouse). This phenomenon is known as MHC restriction^{1,2}. Most evidence now indicates that T cells recognize the antigen and MHC specificities as a single entity via one receptor, rather than having separate receptors for each³⁻⁵.

The receptor molecule is a disulphide-bonded heterodimer of molecular weight 90,000⁶⁻⁹. In the mouse, the two chains are each of approximately 40–45,000 molecular weight, but may be separated on the basis of charge^{6,7}. Tryptic peptide analysis suggests that both the more acidic (α) and more basic (β) chains contain variable and constant regions^{10,11}.

Recently, a complementary DNA clone encoding one chain of the murine T-cell receptor for antigen has been isolated¹². Sequence analysis of this and related clones indicates that the predicted translation products would have significant homology to immunoglobulin¹³ and that like immunoglobulin¹⁴, it consists of variable (V), constant (C) and independently assorting joining (J) regions¹³. It also rearranges in the DNA of different T cells, but not in other cell types¹². The human equivalent of this gene has also been isolated and shown to have some of these same properties²³. N-terminal amino acid sequence analysis of the human T-cell receptor β -chain shows that it represents the product of this gene²⁹. We have studied the components of a rearranged gene isolated from a pigeon cytochrome *c*-specific, H-2E-restricted T_H hybridoma (2B4) and its unrearranged (liver) counterparts¹⁵. Sequence analysis of these genomic clones indicates the presence of a diversity (D) segment in addition to the V, J and C regions^{15,30}, analogous to regions of the immunoglobulin heavy chains¹⁴. It appears that the active gene is formed by juxtaposition of gene segments separated within the genome¹⁵.

There are two distinct C-region genes in the mouse genome which are homologous to the previously described cDNA clones. Each is associated with a cluster of J regions and the whole complex is contained within 15 kilobases (kb). We and others have shown that this gene complex is encoded on chromosome 6 (ref. 16 and M. Kronenberg, personal communication). The 3' cluster of J regions, which includes the 2B4 sequence, has been described elsewhere¹⁵. We present here the sequence analysis of the two C-region genes and the 5' cluster of J regions.

Organization

Genomic clones hybridizing with a constant-region probe were isolated from a B10.A mouse liver DNA library in the vector

λ 1. Their partial characterization has been reported previously¹⁵. Determination of the overall organization of this locus from restriction mapping and sequence analysis (as described in detail below) reveals (Fig. 1) a cluster of J regions preceding a C-region gene in a tandem arrangement: $J_T C_T, J_T' C_T'$. The C_T, C_T' nomenclature has been adopted to avoid confusion between the different J regions and specific exons (that is, C_T1, C_T1', C_T2 , and so on) discussed below. The two C-region genes are separated by about 6.5 kb, each being about 2.5 kb downstream of the respective J-region cluster. Southern analysis shows that all the constant-region segments are contained within this 25-kb segment (data not shown).

Sequence of the J_T cluster

The organization and sequence of the 3' J-region cluster (hereafter referred to as J_T') has been described previously¹⁵. Figure 2a shows the detailed arrangement of the 5' cluster (J_T) and the sequencing strategy, and Fig. 3 gives the nucleotide sequence. The J elements are spaced somewhat further apart in this cluster, up to 440 nucleotides, compared with 40–220 in the 3' cluster¹⁵. J_T contains seven J regions as deduced from sequence homologies with J_T' and immunoglobulin J regions. We have sequenced 2 kb 5' and 1 kb 3' to the regions indicated, without discovering any additional J regions (data not shown). J_T7 is probably a pseudogene because the possible RNA splice sites would make it substantially shorter or longer than the others and may result in out-of-frame splicing to the constant region. Three of the 5' J regions are known to be expressed, as they have been identified from cDNA clones: J_T1 and J_T2 are used in cDNA clone-derived from a concanavalin A-stimulated spleen library (S. M. Hedrick, E. A. Nielsen, J.K. and M.M.D., unpublished results), and J_T3 in the cDNAs 86T1 and 86T5 (see ref. 13). The points of rearrangement to form the active J are indicated in Fig. 3.

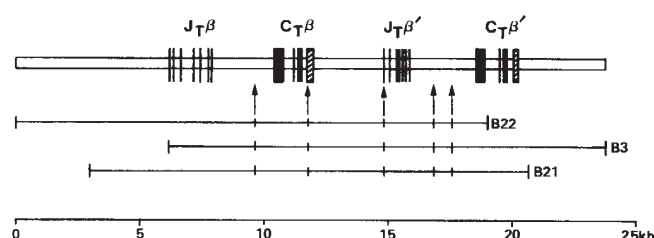


Fig. 1 Sequence organization of the J and C regions. B10.A liver genomic clones were constructed as described previously¹⁵. The arrows represent restriction sites for *EcoRI*. Restriction fragments hybridizing to the mainly constant and J-region probe 86T5 (ref. 13) were subcloned into the plasmid vector pUC9 (ref. 27). Sequencing was performed by the procedures of Maxam and Gilbert²⁸. Exons are represented to scale by solid vertical bars; 3' untranslated regions by hatched bars.

Table 1 Comparison of the domains of the constant region

	External C _T 1	Hinge C _T 2	Transmembrane C _T 3	Cytoplasmic C _T 4	3' UT
Nucleotide changes	1/375	3/18	16/107	4/18	103/217
Amino acid changes	0/125	0/6	3/36	1/6	—
Replacement mutations	0/1	0/3	3/16	1/4	—

Amino acid changes: In C_T3, amino acid 249, Gln: His; 275, Thr: Gly; 278, Val: Leu. In C_T4, 284, Arg: Lys (see Fig. 6 legend). 3' UT, 3'-untranslated region. Deletions/insertions are scored as a maximum of two nucleotide changes.

There are several areas in apparently noncoding regions (Fig. 3) which have short areas of sequence homology with *J* regions. Thus, it seems likely that these represent very degenerate pseudogenes and that the ancestral spacing of the two clusters and even of the immunoglobulin *J* regions may have been similar, one arising from another by duplication.

Figure 4 compares the coding and flanking region sequences of these *J* regions. As in the *J_T*' region and in immunoglobulins, conserved heptamer sequences lie immediately 5' to the *J* regions. Regions similar to the nonamers of the *J_T*' region are also discernible, spaced 12 or 13 nucleotides 5' to the heptamer on the basis of homology to each other. Although maximum homology is obtained with a spacer of 12 nucleotides, as in the *J_T*' cluster¹⁵, the proposed nonamers have more homology to those of the *J_T*' cluster if a spacer of 13 is applied. In the absence of adequate expression data, we have therefore provisionally used a 13-nucleotide spacer. These nonamers are significantly more varied than those of the *J_T*' cluster¹⁵, with only a run of three to five T nucleotides in common. The heptamers are also more varied, although they have the canonical NNNTGTG sequence shared with the *J_T*' cluster¹⁵ and with immunoglobulin heptamers¹⁷. *J_T*1 has the most common of the immunoglobulin heptamers (CACTGTG). This is the only *J_T* element to have this sequence, although it occurs in over 90% of immunoglobulin heptamers¹. There is much variation in the length of the sequence between the heptamer and the expected start of the *J* region. As a result, *J*-region nucleotide sequences can vary in length from 16 to 18 amino acids. In the cDNA clone 86C3, the rearrangement to *J_T*1 occurs so that the asparagine changes to a histidine (Fig. 3 and unpublished data), but in 86T1 and 86T5 the rearrangement occurs just 5' to the serine (Fig. 3 and ref. 13). This produces *J* regions of 15 and 16 amino acids, respectively.

Figure 5 shows the predicted amino acid sequences of these *J* regions and compares them with those from *J_T*' regions. There are several notable differences between the two clusters. In *J_T*6,

one of the glycine residues that is conserved throughout all the other T-cell *J* regions (those that are not obviously pseudogenes) is substituted with an alanine. This sequence, glycine-X-glycine, is found in all immunoglobulin *J* regions except in that encoded by pseudogene *J_K*3, where the first glycine is replaced by a serine^{17,18}. Of the amino acid sequences encoded by the *J_T*' cluster, all but one contains a conserved tyrosine followed by phenylalanine and glycine. In contrast, this tyrosine occurs only twice in the *J_T*-encoded cluster. It is unclear which, if any, of the *J_T* elements are functional other than the three found in cDNA clones.

There is a large block of strong amino acid homology encoded by the 3' end of each *J* region (Fig. 5), with significant differences only between the two glycines and at the 5' end. Thus, diversity is provided by the *J*-region genes in at least four different ways: their large number and high variability (compared with the small number and relative homogeneity in immunoglobulin clusters^{14,18}), their variable lengths and the indeterminate position of rearrangement (P. Patten *et al.*, unpublished).

Structure of constant-region genes

The two C-region genes, C_T and C_T', were sequenced by the strategy shown in Fig. 2b, c. Both genes are arranged in four exons, interrupted by introns at exactly the same positions, the lengths of which are nearly identical. Based either on their sequence similarities to immunoglobulin or the results of Kyte-Doolittle hydropathicity analysis¹³, the four exons can be assigned as encoding the external domain (C_T1), a small hinge-like region (C_T2), a probable transmembrane region (C_T3) and a cytoplasmic tail (C_T4) plus 3'-untranslated region. The nucleotide and predicted amino acid sequences of the two genes are presented in Fig. 6 and the individual exons compared in Table 1.

The sequence 5' to C_T1 (not all of which is shown) indicates that the cDNA clone 86T3 described previously¹³ was formed in part by an 87-base pair (bp) readthrough immediately

Fig. 2 Sequencing strategy for *J_T* cluster (a), C_T (b) and C_T' (c) regions. a Was sequenced from two subclones, *Hind*III to *Bam*HI and *Bam*HI to *Eco*RI. b, Two *Eco*RI subclones were used for sequencing. Because an *Eco*RI site falls within the C_T4 exon, labelling from the vector polylinker *Bam*HI site was used. c, The C_T1' exon was sequenced from a subclone of B22 (Fig. 1). The *Eco*RI and *Pst*I sequencing from the wavy line indicate where this subclone ends. The other exons were sequenced from a subclone of B3 which extends further on the 3' side. Sequencing of the subclones was by the procedure of Maxam and Gilbert²⁸ by either 5'-end labelling with polynucleotide kinase (thin lines) or 3'-end labelling with the Klenow fragment of DNA polymerase I (thick lines). The broken line in b indicates one fragment sequenced by the ³⁵S-dideoxy method.

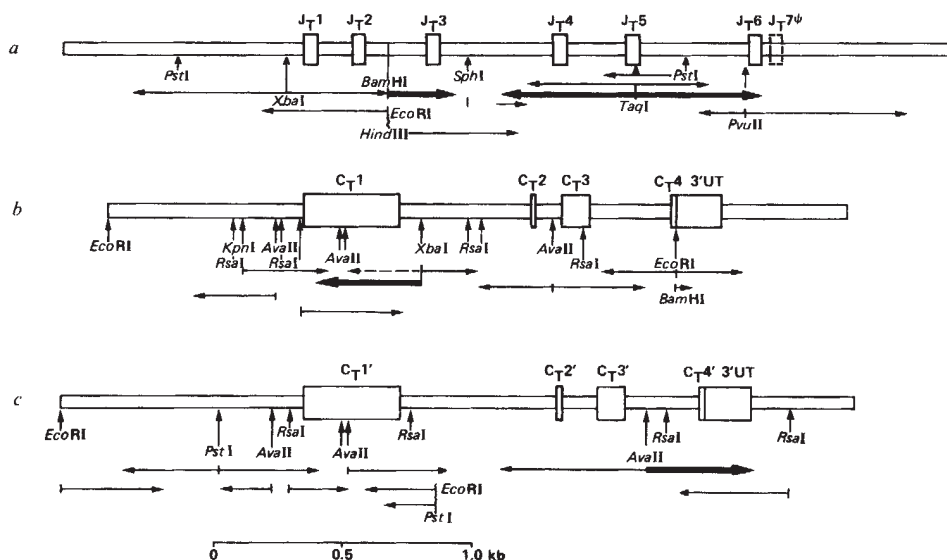


Fig. 3 Sequence of the genomic J_T regions obtained by the strategy of Fig. 2. Translations of the J regions are shown. That of the clear pseudogene (J_{T7}) is shown to be the same length as the apparently functional J regions even though its splice site is displaced (see Fig. 4). The deduced positions of rearrangement of the four cDNA clones [86C1, 86C3 (unpublished), 86T1 and 86T5 (ref. 13)] are indicated. It is not possible to determine the exact position for 86T1 and 86T5, so the two possible positions are indicated.

		J_{T1}	
TCTAGATACTAGCTGGGAAGAGCCTCTCTCCACCCCTTAAGATTATTTCTCCTCATCTATGCGACGTGCAACACAGCAAGCTCTCTTTGGTAAAGAACACGACTCACAGTTGTAG		AsnThrGluValPhePheGlyIleGlyThrArgLeuThrValVal	120
		86C3→	
GTAAGATATCTTTAGGTAATTTAGGTCCTCTCTCGGACAAAGCATGCTCCGTGTCATATTCAGATCTGTATCTGTGTGCAACATCCGACTACACCTTCGGCTCAGGAGC		AsnSerAspTyrThrPheGlyIleSerGlyIleThr	240
		86C1→	
ArgLeuLeuValIle			
AGGCTTTTGGTAATAGGCTAGGGCTCTTGGGTGTGGATGGAAAGTGAATACTGTTCTGGTAATCAGAGGAAGATGCTGCTCATGGTTCTCTTGAGGAAGAGGCTCTCT			360
CATGTCTGGGACTACATTTCTAATGAGGCTGGATCCGCGAGTCGACGTGATCCACAAGGTGGATTCATCATAGGCTAGGAATGTTATGGTGGAGTCAAGATGAACCTCGGGTGAAG			480
		J_{T3}	
TCAGGAAGAGAGAGGAGAAAGGCTCACAGAGAGGAGGGGTTTTGAAGTGGACCCGGGAGGCTGTGTTCTGGAAATACGCTCTATTTGGAGAAGGAAGCCGCTCATTTGTTGAGGT		SerGlyAsnThrLeuTyrPheGlyIleGlySerArgLeuIleValVal	600
AAGTTAGGCGCAATAGGCTGGTACTGGGTGAAGTTTCTCTGAGGTGCTAGGTTGCGAGTACTCTATCAGAAAGAACTAACTCCTCTGCTCATGTCATGACAGGAAGTGAATG		86T1/T5→	720
CCCACAGGTGGCTATGCGAACTCGGCATCGTTTAAATGAGGCAAGCTCAACTGGGAGTAAAGCTCATGTACTGTTACAGTCTCAGGTTTGGCTCTCCATATAGGCAGAGACAGGC			840
TTGGCAGGGAAGATGTATGGCCATTAGGAACGTGATGTCTAATTACATTAGTAATATGGCATAGTTGTCTGTCTCTCTCTCCAGAGATAACACCTTCTCAGGAATGAGGAA			960
		J_{T4}	
GCTAGCTGCTAGGCAGGCACAGGAGGAGGGGAGGAGCATGAGGAAGTTTACCCTAGGCTTAAATGTTGTTTCCACGAAGATTATTTTCGGCTATGGAACCAAGCTGTCTG		SerAsnGluArgLeuPhePheGlyIleGlyThrIleLeuSerVal	1080
a1Leu			
TCTTGGGTATGTAAGATTCTTTCTCGGAGGGTGTGTTGTAACCTAGAGCCACGGGAACTCTCGGCAAGGTGAGGACTGCAAGGCAGCTTGGCTGTGGTTTCAGGAGGAAG			1200
GAAGGATGGTGAATGTGCTGAGCCCTTGGAGATCTGTGACAGTCTGCTCAAGAAGGGTCCCATCAGCTCTTTGGAGTTTGTACCCCTCATGTTGACTGTGTAAACCAAGCT		GlnProAla	1320
		J_{T5}	
ProLeuPheGlyIleGlyThrArgLeuSerValLeu			
CCGCTTTTGGAGGGGACTCGACTCTCTGTTCTAGTAACTATGGACCAAACTGGTGGACCAATTGCTTTGGACCTGGAGTCTGTAAACCCAGAGCTGGTGAAGGAGCTGG			1440
GAAATTTAGTTATCTTTACTTGCTAGGTTCTGAAATGGAACCTGCTCACTATGAGGAGATGCATCAGCACCTGCTTTTCCCTGAGCTGCAAGCTTAGGAGCTTTTCTCTGC			1560
CTTCTCCCAAGGTATCTACAGTTTGTGTCATGGGTGCTCTCTGTGAGTCCCTTCCCTGCTCTCTCTGTGACATGCTCATGTTATTACTGGAAGCTGGAATACTCTCTG			1680
CAGGCAGAGAGGTAGTCATTAGGCTCTCAGGATCTAAGACTCGCTCTAGCTTTGGGACCGCTGGTAGGTTTACCAAGGTACCTGCAGCTGTGTTCTTATAATTCGCCCTCTACT		SerTyrAsnSerProLeuTyrP	1800
		J_{T6}	
heAlaAlaGlyIleThrArgLeuThrValThr			
TTGCGCAGGCACCGGCTCACTGTGACAGGTATGGGGCTCCATTTCTGACTGGAGGGGGCTGTGCTGTGTTGGATGACCTGCTTGGAAAGGAACCTAGGTATAAGATAAGACAG		ProValLeuAspAspHisGlyLeuGlyIleGlyIleLeuArgTyrLys	1920
AATCCTTAGGTATAAGTAAGACAGAGTCTGCTCTCCATCTGGAACCTCTACCTTTCTACCTTCCCGAGGCTGGTGGCTTGTCTCTCTGGTTTGTCTGTCTGTTCTGACAC			2040
AGAATGCTTTCTCAGGGCTTGGAGTCTTCTGCTCTCTTTTACTATTCTTTGGGCTACTGTCTCAGTGACTCTGAGCGGATGACTTCTGTTTCTCTAGCCTCATCTCC			2160
		J_{T7}	

upstream of the C region. The sequence 5' to this section of the cDNA is derived from another region.

The two C -region genes are strikingly similar over their entire coding-region sequence: they produce a total of only 4 amino acid substitutions out of 173. The positioning of the nucleotide and amino acid changes is particularly intriguing, as they are limited almost exclusively to the 3' exons. Only one (silent) nucleotide change occurs within the external (C_{T1}) domain, and three silent mutations in the 18 nucleotide stretch of C_{T2} . The sequences encoded by the transmembrane (C_{T3}) and cytoplasmic (C_{T4}) exons do show some amino acid changes (Table 1), but most of these changes would probably not markedly affect the proteins produced, as threonine and glycine (which occur at position 275) are both uncharged polar amino acids, valine and leucine (278) are both non-polar and arginine and lysine (284) both basic. (This amino acid numbering refers to ref. 15 which supercedes that of ref. 13.) Of the four amino acid replacements, only the glutamine for histidine at position 249 is likely to be an allelic or allotypic determinant as it is the only non-conservative change (glutamine is uncharged, histidine often basic) and possibly external to the cell membrane (because it is at the 5' end of the transmembrane exon, C_{T3}). The 3'-untranslated regions are almost completely divergent (~45%

homologous) and differ somewhat in size (216 nucleotides in C_T and 174 in C_T'). The overall divergence of the intron sequences is similar to that of the 3'-untranslated region except at the boundaries with the coding-region sequence. Assuming that noncoding DNA diverges at a rate of 1% every 2.2 Myr²⁰, this indicates that the two genes separated at least 120 Myr ago. The stronger conservation of more 5' regions has also been seen in immunoglobulin genes²¹, although to a much lesser extent than observed here.

Clearly, some very strong mechanism has been operating to maintain the similarity between the two copies of each exon of the C -region genes. The data argue against a recent gene conversion event occurring over the whole region of the gene. However, it is likely that gene conversion has occurred in the first exon fairly recently because it has almost complete nucleotide sequence conservation, and, given that the amino acid changes in the other exons are mainly silent, it is more likely that homology in these regions is maintained by selection at the protein level rather than by gene conversion. The close nucleotide and amino acid homology to the human equivalent²² indicates that there must have been strong functional constraints placed on this gene in its recent evolution, much more than for immunoglobulins¹⁴.

Fig. 4 Flanking sequence homologies between J regions of the J_T cluster. The sequences are split to show (from left to right) nonamers, spacers (of 13 nucleotides in the apparently functional J regions), the heptamers, coding regions and presumed splice sites. Homologies in flanking sequences are underlined. The nonamers have been arranged for consistency with the J_T' cluster, but more homology would be obtained by positioning them one nucleotide to the right.

9	7	J_T CODING REGIONS	
TATTTTCT	CCTCATCTATGG	CAACACAGAAGTCTCTTTGGTAAAGAACAGACTCACAGTTGTAG	J_{T1}
CCATATTCG	AGTATCTGTATTC	CAAACTCCGACTACACCTTCGGCTCAGGAGCAGGCTTTTGGTAATAG	J_{T2}
GGGTTTTGA	AGTGAACCCGGGA	TTCTGGAATACGCTCTATTTTGGAGAAGGAAGCCGCTCATTTGTTGTAG	J_{T3}
AAGTTTTAC	CACTAGGCTTTAA	TTTCCAACGAAGATTATTTTTCGGTCAATGGAACCAAGCTGTCTGTCTTGG	J_{T4}
GGAGTTTGT	CACCTCATGTTG	TAAACACAGCTCCGCTTTTGGAGAGGGGACTCGACTCTCTGTTCTAG	J_{T5}
AGGTTTTAC	CAAGGTCACTGC	TTCTATAATTCGCCCTCTACTTTGCGCAGGACCCGCTCACTGTGACAG	J_{T6}
GCTCCATT	CTGACTGGAGGG	CCTGTGTTGGATGACCATGGCTTGGAAAGGAACCTTAG	J_{T7}

Fig. 5 Amino acid sequence homologies (outlined) between the seven J_T J regions compared with those of the six apparently functional J regions of the J_T' cluster¹⁵. The J_T -7 pseudogene is shown to end at the same position as the apparently functional J regions (see Fig. 4 legend).

The sequence encoded by the C_T3 exon was predicted to be the transmembrane portion of the molecule because of the hydrophobicity of the amino acids in this region¹³. Comparison of the amino acid sequence encoded by this segment with that of the transmembrane regions of the various immunoglobulin heavy-chain isotypes shows that although the immunoglobulin transmembrane regions have a high degree of homology to each other²⁵, they bear little resemblance to that of the T-cell receptor, except inasmuch as the amino acids are uncharged. This different transmembrane sequence, including the anomalous lysine at position 265, may represent functional requirements within the membrane which are involved in the association with the other components of the receptor complex, such as the α -chain of the receptor or the mouse T3 equivalent.

Fig. 6 Sequence of the four constant-region exons (with flanking sequence) from C_T (upper) and C_T' (lower). The predicted amino acid sequences are shown in italics. Only the C_T sequences are shown in full, with the differences found in C_T' indicated below. Homology is shown by a dashed line. Spaces have been inserted into flanking sequences to maximize homology. The 3'-untranslated region following C_T4 contains a poly(A) addition signal, AATAAA, which is boxed. The position of the poly(A) addition site is also indicated. It is known from the sequence of the cDNA clone from 2B4 (ref. 15) that this is correctly assigned in C_T' ; however, that of C_T is only deduced from the homology to C_T' . Note that the amino acid numbering system follows that of ref. 15 and supercedes that of ref. 13.

Discussion

We have isolated two closely linked constant-region genes that cross-hybridize to the cDNA clones described previously¹³. Each *C* region is associated with a set of *J* regions which account for all of the nine *J*-region elements so far identified from cDNA clones. In the cDNA clones analysed thus far, we have found that a given *J_T* element is always expressed with the constant region immediately downstream. Together, these two clusters may encode as many as 12 functional *J* elements, much more than any immunoglobulin cluster (usually containing four to six minigenes)^{14,18}.

The common evolutionary origin of the genes for immunoglobulins and the T-cell receptor has left the T-cell receptor β -chain gene with many features found in either the heavy- or light-chain immunoglobulin genes. The spacing between the *J_T* and *C_T* elements is more like that between light chains than between *J_H* and *C_H* (ref. 14). The genomic organization, *J-C/J-C*, is like that of λ light-chain genes, yet the *D* region¹⁵, the separate hinge-like exon for the cysteine that we believe forms the interchain disulphide bond¹³, and the presence of transmembrane region and cytoplasmic tail exons are all characteristic of immunoglobulin heavy-chain genes. The length of the T-cell *J* region is closer to that of heavy-chain *J* regions than it is to those of light chains (ref. 15 and this paper). It has already been noted that the nucleotide sequence homologies are intermediate between those of heavy- and light-chain genes¹³, yet the sequence of *C_T*1 is much more like the light chain²² than is the rest of the gene (40% compared with 25%).

The apparent signal sequences for *V-D-J* joining are remarkably like those of immunoglobulin. This and the existence of the *D* region (as in heavy chains) indicates that the basic

mechanism of *V-J* and *V-D-J* joining is similar in both systems and that they have a common evolutionary origin. Because cellular immunity is thought to occur much earlier in phylogeny than humoral immunity²⁶, it seems reasonable to speculate that the *V-D-J* joining mechanism of immunoglobulin genes arose directly from that of the T-cell receptor genes studied here.

Why are there two essentially identical *C*-region genes? The expression of the two does not seem to correlate with functional isotypes, as both *T_H* and *T_C* cells use either locus (unpublished results). Perhaps the answer lies in the diversity of the *J*-region segments. Both here and in immunoglobulin genes (reviewed in refs 14, 18), the *J* clusters contain very few (4–6) apparently functional minigenes. There may be a limit to the number of these elements that can be clustered, so that if, for example, T-cell diversity requires 10 *J* regions, they would have to split into two clusters, perhaps to avoid aberrant DNA rearrangement or RNA splicing. Although there is no clear correlation between function and the expression of a particular constant region, we have found 30-fold higher expression of *C_T* than *C_T'* in BALB/c thymocyte and Concanavalin A-activated spleen cDNA libraries. As most cDNA clones derived from functional T-cell lines and hybrids use *C_T'*, the significance of this is uncertain. An explanation for the presence of the two *C* regions based on maturation cannot be discounted. We are currently testing this and other possibilities.

We thank Kathy Redman for help in the preparation of this manuscript and J.C.-K. Shen for his constructive criticisms. N.R.J.G. is a fellow of the Cancer Research Institute of New York and D.M.B. is supported by the Stanford Alumni medical scholars programme. This work was supported by grants to M.M.D. from the NIH (AI19512 and AI20320) and the American Cancer Society (IM00364).

Received 22 May; accepted 8 June 1984.

1. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 52–177 (1979).
2. Sprent, J., Korngold, R. & Molnar-Kimber, K. *Springer Semin. Immunopath.* **3**, 213–245 (1980).
3. Kappler, J. W., Skidmore, B., White, J. & Marrack, P. *J. exp. Med.* **153**, 1198–1214 (1981).
4. Hunig, T. R. & Bevan, M. J. *J. exp. Med.* **155**, 111–125 (1982).
5. Heber-Katz, E. *et al. J. exp. Med.* **155**, 1086–1099 (1982).
6. Haskins, K. *et al. J. exp. Med.* **157**, 1149–1169 (1983).
7. Samelson, L. E., Germain, R. N. & Schwartz, R. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6972–6976 (1983).
8. Allison, J. P., MacIntyre, B. W. & Block, D. J. *Immun.* **129**, 2293–2300 (1982).
9. Meur, S. *et al. J. exp. Med.* **157**, 705–719 (1983).
10. McIntyre, B. W. & Allison, J. P. *Cell* **34**, 739–746 (1983).
11. Kappler, J. W. *et al. Cell* **35**, 295–302 (1983).
12. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. *Nature* **308**, 149–153 (1984).
13. Hedrick, S. M., Nielsen, E. A., Kavaler, J. Cohen, D. I. & Davis, M. M. *Nature* **308**, 153–158 (1984).

14. Honjo, T. *A. Rev. Immun.* **1**, 499–528 (1983).
15. Chien, Y., Gascoigne, N. R. J., Kavaler, J., Lee, N. E. & Davis, M. M. *Nature* **309**, 322–326 (1984).
16. Lee, N. E. *et al. J. exp. Med.* (in the press).
17. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
18. Tonegawa, S. *et al. Cold Spring Harb. Symp. quant. Biol.* **45**, 839–858 (1980).
19. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. *Sequences of Proteins of Immunological Interest* (U.S. Dept of Health and Human Services, Washington DC, 1983).
20. Zimmer, E. A. thesis, Univ. California, Berkeley (1981).
21. Ollo, R. & Rougeon, F. *Cell* **32**, 515–523 (1983).
22. Yanagi, Y. *et al. Nature* **308**, 145–149 (1984).
23. Honjo, T. *et al. Cell* **18**, 559–568 (1979).
24. Oi, V. T. *et al. Nature* **307**, 136–140 (1984).
25. Komaromy, M. *et al. Nucleic Acids Res.* **11**, 6775–6785 (1983).
26. Manning, M. J. & Turner, R. J. *Comparative Immunobiology* (Wiley, New York, 1976).
27. Viera, J. & Messing, J. *Gene* **19**, 259–268 (1982).
28. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
29. Acuto, O. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 3851–3855 (1984).
30. Kavaler, J., Davis, M. M. & Chien, Yueh-hsiu *Nature* **310**, 421–423 (1984).

LETTERS TO NATURE

Microwave anisotropy due to cosmic strings

Nick Kaiser

Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA

Albert Stebbins

Department of Physics, University of California, Berkeley, California 94720, USA

There has been much interest in the cosmological consequences of line-like topological defects ('cosmic strings') which may be produced in a phase transition in the early Universe^{1–6}. Such strings would produce density fluctuations on a broad range of scales and may have been responsible for the formation of galaxies. In view of this possibility, it would be useful to have more direct evidence of the existence of strings than the observed clustering of matter.

We show here that cosmic strings would generate anisotropy of the microwave background temperature with a very distinctive feature; the temperature would have steplike discontinuities on curves on the sky. Present anisotropy measurements limit the line density parameter, $G\mu$, to be $<10^{-5}$. Popular values of this parameter, for the purpose of galaxy formation, are only an order of magnitude below this. The anisotropy that we discuss will exist in addition to that produced by curvature fluctuations due to strings and matter at the time that the Universe becomes optically thin.

Cosmic strings are configurations of the matter fields which owe their stability to the topology of the space of degenerate vacua produced in a phase transition in the early Universe. For our purposes, we can ignore the internal structure of the strings and treat them as one-dimensional objects with tension. In the rest frame of the string, the mass per unit length, μ , is equal to the tension (we use units where the speed of light, $c = 1$). The equality of the line density and the tension causes the typical transverse velocity associated with large vibrations of the strings to be close to the speed of light. The strings that we consider

cannot end (certain types of strings might end in monopole-like objects), but must either close on themselves or extend to infinity. Closed strings will be called loops.

The dynamics of cosmic strings have been widely discussed¹⁻⁷. As with any object in tension, strings will accelerate so as to try to become straight. Everett³ has shown that damping of string motion due to their non-gravitational interaction with other matter becomes negligible soon after the strings form. Strings which extend outside the horizon will be conformally stretched by the cosmic expansion while any wiggles on scales smaller than the horizon are damped on a time scale comparable with the expansion time. Thus, at a given epoch these strings will be straight on length scales smaller than a horizon size but will be quite convoluted on length scales larger than this. The typical velocity associated with the straightening of a string will be close to the speed of light as noted above. Thus, the velocity field of a string extending outside the horizon will be relativistic and approximately constant over scales much smaller than a horizon size. Once a loop enters the horizon it will no longer expand, but will rather oscillate with a period comparable to the light travel time across it. This motion will be damped by gravitational radiation, causing the size and period of the loop to decrease approximately linearly with time. The fractional decrease in size, period and mass of the string in one oscillation is given approximately by $G\mu$ where G is the gravitational constant. A string will decrease to zero size in a finite amount of time, losing its energy by gravitational radiation. The ultimate fate of loops will be determined by quantum effects when the loop is microscopic in size and will have no cosmological significance.

The distribution of strings in our Universe is not quite so well understood. After the phase transition the strings will form a random network of self-avoiding curves. Some of this string will be in closed loops and some will be in infinite strings. The fraction of length of string in both should be of the order of unity. If the initial mean separation of strings is of the order of the horizon size, as seems likely, then for $G\mu \ll 1$ the strings contribute a fraction $\sim G\mu$ to the total density. Turok⁶ has argued that the initial distribution of closed loops is self-similar in the sense that the fraction of space filled by loops of size $\sim L$ is independent of L . Such a distribution gives a constant number of loops entering the horizon per horizon volume per expansion time. If the infinite strings simply straighten out, then the number of open strings across a horizon-sized volume increases with time and the strings would soon come to dominate the density⁷. However, provided intercommutation (exchange of partners when strings intersect) occurs, the number of open strings across a horizon volume remains of the order of unity, with roughly one horizon-sized closed loop being created per horizon volume per expansion time. These loops would be formed with relativistic net velocities. This net velocity will then decay as $(1+z)$ just as with massive particles where z is the redshift.

Birkinshaw and Gull⁹ have discussed how to detect transverse velocities of clusters of galaxies by observing the temperature deviations of the microwave background in the surrounding region of sky. The radiation which passed in front of the cluster will be made redder, and that which passed behind the cluster will be made bluer. This is much the same effect as is used by operators of interplanetary spacecraft who may increase (decrease) the energy of the spacecraft in the Sun's rest frame by sending them closely behind (in front of) an orbiting planet. The magnitude of the fractional temperature deviation is given by $\delta T/T \approx \beta\phi$ where β is the transverse velocity in units of c , and ϕ is the angular deflection produced by the moving object in the frame of the object. The temperature fluctuation is small in the case of clusters because the peculiar velocity of clusters is much smaller than the speed of light. As noted above, strings will naturally move at velocities close to the speed of light. However, the results of Birkinshaw and Gull are not directly applicable to strings as the gravitational fields of strings are different from those of ordinary matter.

Vilenkin² has shown that the geometry produced by the gravitational field near a length of straight string is that of Minkowski

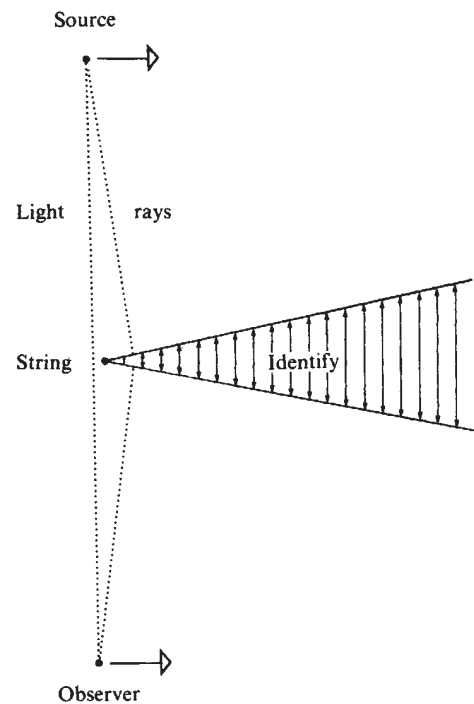


Fig. 1 The geometry of space-time near a string. A cross-section of space-time perpendicular to a length of string as seen from the rest frame of the string. The space-time is locally Minkowskian but a wedge has been removed and the exposed faces identified. According to an observer a string is between him and a source, and is moving to the left. He sees two images of the source. According to the space through which the left light ray passes, the observer is at rest with respect to the source and there is thus no Doppler shift. According to the space through which the right ray passes, the observer is moving towards the source. Thus the observer's right image of the source is blueshifted with respect to the left image. All light rays passing to the right of the string will be blueshifted and all light rays passing to the left will experience no Doppler shift. Similarly, a string passing between us and the 'cosmological photosphere' will cause a discontinuity in the temperature of the microwave background along its length where it is moving transversely to our line-of-sight.

space with a three-dimensional wedge taken out of each space-like slice. The vertex of the wedge lies along the length of string and the angle subtended by the missing wedge in the rest frame of the string is $8\pi G\mu$. The two 'exposed' faces are identified. Thus space-time remains flat everywhere except along the string where it is highly curved. A cross-section of the rest frame, perpendicular to the string, is given in Fig. 1.

If $G\mu \ll 1$, then the stress-energy of the string produces only small (linear) perturbations from the metric of the rest of the Universe. Thus, we may add up the gravitational fields from the different sources just as one does in Newtonian gravity. Because the matter in the Universe does not produce significant perturbations from the Minkowski metric on scales less than the horizon, it is clear that the gravitational field, at a point much closer to a length of string than both the horizon size and the radius of curvature of the string, will essentially be the same as the gravitational field at a similarly located point in Minkowski space. In other words, the geometry very near cosmological strings is the same geometry which is depicted in Fig. 1. The relative velocities of two objects in this space are ambiguous, depending on which way one parallel transports their velocities around the string. Because the space in which any particular photon travels is a piece of Minkowski space, the usual Doppler formulae apply. Consider a length of string moving with relativistic velocity perpendicular to our line-of-sight and two nearly parallel light rays reaching an observer which have passed very close to, but on opposite sides, of the string (see Fig. 1 but note

that strings also produce double images). We see that the ambiguity in relative velocities results in a discontinuity in the Doppler shift of a source behind the string. Thus, there will be discontinuities in the microwave background temperature at positions on the sky where strings with relativistic transverse velocities are located. It is this discontinuity in temperature along curves in the sky which is peculiar to strings. This effect has been noted independently by Gott¹¹.

In the rest frame of the string all particles passing the string are deflected by an angle of $8\pi G\mu$ with respect to particles passing on the other side of the string. Unlike usual gravitational lenses, this deflection is independent of impact parameter as long as the impact parameter is much smaller than the radius of curvature of the string. Thus the magnitude of the discontinuity in temperature across the string is $\delta T/T = 8\pi G\mu\beta$ where β is the transverse velocity of the string which will typically be close to unity. This jump in temperature will persist an angular distance away from the string corresponding to the apparent angular size of the radius of curvature of the string. The magnitude of the temperature jump is independent of the redshift at which the light rays reaching us passed by the string.

The number density of discontinuities on the sky depends on the number of loops appearing on the horizon per expansion time which is determined by the efficiency of intercommutation. However, it is difficult to see how this process could cause the number of horizon size lengths of string per horizon volume to decrease below unity. We thus consider one horizon size length of string per horizon size to be the minimal model for the production of anisotropy.

We wish to calculate the general properties of the microwave sky anisotropy in this minimal model. Let us assume that the microwave photons were last scattered at redshift z_{ls} . In a perfectly homogeneous universe, the matter becomes mostly neutral and optically thin at $z \sim 1,000$. However, in a universe with strings, there will be large-amplitude inhomogeneity on small scales and the heat input from objects forming at or before z_{rec} may reionize the plasma⁸. If the plasma were kept fully ionized then $z_{ls} > 10$, so we must have $1,000 > z_{ls} > 10$. The angle subtended by a horizon-sized volume of space at z_{ls} is $\theta_{ls} \sim z_{ls}^{-1/2} \ll 1$. What do we expect to see on a round patch of sky of diameter $\theta > z_{ls}^{-1/2}$? It is clear that the one horizon size length of string per horizon volume at redshift z will project to one length of string of angular size θ if $z < z_{ls}$. These strings will be moving relativistically as they were unable to straighten themselves out on these length scales before this epoch but will be smooth on scales smaller than this. Thus these strings will produce line-like discontinuities in temperature with $\delta T/T \sim 8\pi G\mu$ which persist for an angle $\sim \theta$ away from the line.

Also projected on the sky will be relativistically oscillating loops which have already entered the horizon. However, in our minimal model one can show that the angular distance between such sub-horizon loops is much greater than their angular size. Thus, in a beam switching experiment with beam width and beam throw $\sim \theta$ it is unlikely that the beam throw would straddle a sub-horizon loop, but if $\theta > \theta_{ls}$ it would probably straddle a horizon size loop. Such an experiment would average over any anisotropy due to loops with projected size smaller than θ and would only be sensitive to strings with projected angle of curvature $\sim \theta$. For the minimal model we conclude that a beam switching experiment on angular scales $\theta > z_{ls}^{-1/2}$ rad would see temperature deviations with magnitude $\sim 8\pi G\mu$, due mostly to horizon-sized lengths of string. As such experiments have already been done which do not detect anisotropy of magnitude 10^{-4} (ref. 10) on angular scales $> 10^\circ$, we conclude that $G\mu < 10^{-5}$ if any strings exist. For purposes of galaxy formation, $G\mu \approx 10^{-6}$ has been suggested⁷. If more strings exist than in our minimal model then an even greater anisotropy of $\sim \sqrt{N_s} 8\pi G\mu$ would be produced, where N_s is the mean number of overlapping loops of a given size. If anisotropy is ever detected, one could determine whether it was due to strings by looking for line-like temperature discontinuities on the sky. From the magnitude of the discontinuities one can read off the microphysical parameter

μ . If strings were found in this way it would test the theory of general relativity for stress-energy tensors unlike those of ordinary matter, as well as confirm some ideas about non-perturbative quantum field theories.

Received 23 April; accepted 24 May 1984.

1. Kibble, T. W. B. *J. Phys. A* **9**, 1387–1398 (1976).
2. Vilenkin, A. *Phys. Rev. D* **23**, 852–857 (1981).
3. Everett, A. E. *Phys. Rev. D* **24**, 858–868 (1981).
4. Vilenkin, A. *Phys. Rev. D* **24**, 2082–2089 (1981).
5. Vilenkin, A. & Shafi, Q. *Phys. Rev. Lett.* **51**, 1716–1719 (1983).
6. Turok, N. Preprint, *Grand Unified Strings and Galaxy Formation* (Univ. California, Santa Barbara, 1984).
7. Turok, N. Preprint, *Stretching Cosmic Strings* (Univ. California, Santa Barbara, 1984).
8. Hogan, C. Preprint, *Massive Black Holes Generated by Cosmic Strings* (Caltech, 1984).
9. Birkinshaw, M. & Gull, S. F. *Nature* **302**, 315–317 (1983).
10. Fixsen, D. J., Cheng, E. S. & Wilkinson, D. T. *Phys. Rev. Lett.* **50**, 620–622 (1983).
11. Gott, J. R. Preprint, *Gravitational Lensing Effects of Vacuum Strings: Exact Solutions* (Princeton Univ., 1984).

'Melting ice' I at 77 K and 10 kbar: a new method of making amorphous solids

O. Mishima, L. D. Calvert & E. Whalley

Division of Chemistry, National Research Council, Ottawa, Canada K1A 0R9

Amorphous solids are made mainly by cooling the liquid below the glass transition without crystallizing it, a method used since before recorded history¹, and by depositing the vapour onto a cold plate², as well as by several other methods^{3,4}. We report here a new way—by 'melting' a solid by pressure below the glass transition of the liquid—and apply it to making a new kind of amorphous ice. Thus, ice I has been transformed to an amorphous phase, as determined by X-ray diffraction, by pressurizing it at 77 K to its extrapolated melting point of 10 kbar. At the melting point, the fluid is well below its glass transition. On heating at a rate of $\sim 2.6 \text{ K min}^{-1}$ at zero pressure it transforms at $\sim 117 \text{ K}$ to a second amorphous phase with a heat evolution of $42 \pm 8 \text{ J g}^{-1}$, and at $\sim 152 \text{ K}$ further transforms to ice I with a heat evolution of $92 \pm 15 \text{ J g}^{-1}$. In one sample, ice Ic was formed and in another, existing crystals of ice Ih grew from the amorphous phase. Heating below the 117 K transition causes irreversible changes in the diffraction pattern, and a continuous range of amorphous phases can be made. Similar transformations will probably occur in all solids whose melting point decreases with increasing pressure if they can be cooled sufficiently for a transformation to a crystalline solid to be too slow.

Several solids, such as ice I, melt with a decrease of volume, and so the melting temperature falls as the pressure increases. The melting line usually ends at a triple point, where another solid phase becomes stable, but, as the solid-liquid transition is first-order, it can, in principle, be extrapolated to low temperature and even to zero temperature. It follows that if such a solid is compressed at a temperature that is low enough to prevent transformation to another crystalline phase, it must either 'melt', perhaps to an amorphous solid if it is below the glass transition of the liquid, or become a crystal that is greatly superheated into the liquid region. Either case would be new and could be of great interest. The melting curve of ice Ih extrapolates to $\sim 10 \text{ kbar}$ at 77 K (Fig. 1); we have therefore squeezed ice Ih at 77 K and have examined the product that is recovered at zero pressure by determining its density, by thermal analysis, and by X-ray diffraction.

Figure 2 plots the compression for four different runs against the nominal pressure, assuming no friction. In each run, a transition started at an estimated real pressure of $10 \pm 1 \text{ kbar}$ and appeared complete by $\sim 15 \text{ kbar}$. About two-thirds of the volume change occurred in a pressure range of $\sim 0.7 \text{ kbar}$. The transition goes surprisingly easily for such a low temperature. For comparison, a similar sample of ice IX at the same temperature did not transform below 25 kbar, although it is always metastable relative to ice II and also becomes metastable relative to ice VI

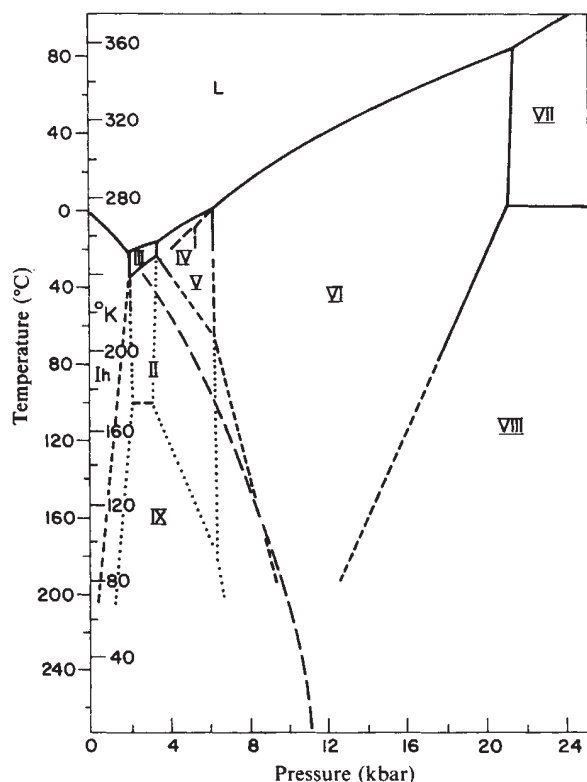


Fig. 1 The phase diagram of ice in the pressure-temperature plane. The melting line of ice I is extrapolated as a dashed line.

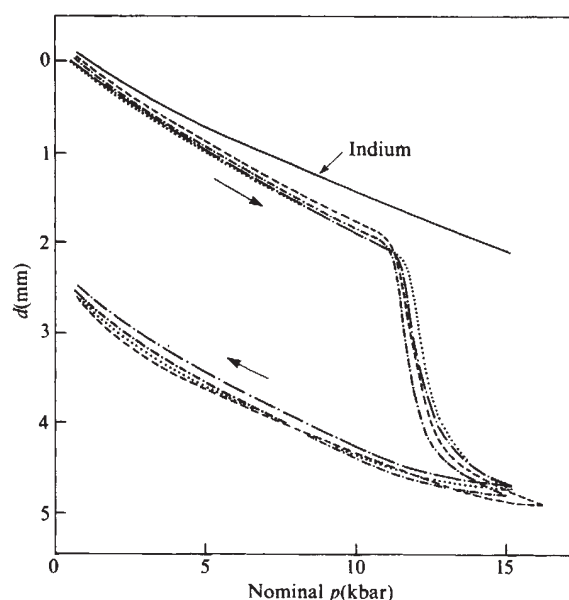


Fig. 2 Compression of ice I as a function of the nominal pressure on the sample for four independent runs at 77 K and on a volume of indium equal to the volume of the indium cup and the ice. About 1.2 cm³ of water in an indium cup was compressed in a steel cylinder by a hydraulic press and the displacement d of the piston was measured to 2.5 μ m by a dial gauge as a function of the nominal pressure p . The pressure of the sample is ~ 0.9 of the nominal pressure. The measured compression was used with the compression of an equal volume of indium¹⁷ to determine the compression of the ice.

at ~ 9.4 kbar and to ice VIII at ~ 12.3 kbar. The ease with which ice I transforms at ~ 10 kbar is, therefore, quite unprecedented. The specific volumes of liquid water, ice and the new phase as a function of pressure are represented in Fig. 3, which may be read in conjunction with Fig. 1. The specific volume at 77 K of ice I and of the new phase, both at 10 kbar, are consistent with the extrapolated specific volumes of ice I on the liquid-I line (labelled a), of the liquid on the liquid-I line (labelled c) and of the liquid along the 9.8-kbar isobar (labelled d) where the dotted lines are extrapolations. The relation of the new phase to the amorphous ice made by quenching the liquid can be understood by the extrapolation of the liquid along the 1 bar isobar (labelled b).

Figure 4 shows a first-order plot of the heating curve of the recovered phase. There are exothermic events starting at ~ 117 and ~ 152 K, at a heating rate of ~ 2.6 K min⁻¹, with a heat evolution, from two independent experiments, of 42 ± 8 and 92 ± 15 J g⁻¹, respectively.

Two specimens were analysed by X-ray powder diffraction at ~ 95 K⁵, and microphotometer traces of representative patterns of the second are reproduced in Fig. 5. Both specimens as recovered had a halo pattern typical of an amorphous material, having its main peak at 3.0 Å and with rings and spots ascribable to a small amount of untransformed ice Ih. After heating the first specimen, which was a powder, to ~ 130 and ~ 170 K and cooling to 95 K, the halo pattern was replaced by a ring pattern typical of ice Ic, and the ice Ih pattern remained. After heating to ~ 200 K, only the ice Ih pattern remained, and the original Ih crystals had clearly grown.

The second specimen was a euhedral fragment ~ 0.4 mm on edge with a small amount of powder. The principal halo was a broad peak at 3.0 Å; it narrowed and its position, as measured at ~ 95 K, shifted almost linearly with heating temperature, from 3.0 to 3.65 Å after heating successively to ~ 105 , ~ 112 and ~ 135 K for ~ 10 min and further shifted to 3.67 Å after heating to 155 K. After heating to ~ 175 K for ~ 20 min, the halo pattern disappeared and the Ih spots grew considerably, but no Ic

pattern was produced. This appears to be the first report of the direct transformation of amorphous ice to ice Ih instead of Ic. The conditions no doubt gave preference to growth of Ih crystals against nucleation of Ic.

The first halo pattern is undoubtedly of a new dense amorphous solid having its principal peak at ~ 3.0 Å; this is much smaller than that of the first ice-Ih triplet centred at 3.65 Å or the principal peak of amorphous ice made by condensing the vapour⁶⁻⁹ or rapidly cooling liquid water¹⁰. The bands are rather broad, showing that the interatomic correlations are relatively weak. When this phase is heated successively to ~ 105 , ~ 112 and ~ 135 K, the halo moves to longer spacings, the bands become narrower and a discontinuous change occurs at ~ 117 K, as shown by the heating curves. The discontinuous change is probably caused by 'runaway' heating. The phase so produced resembles the amorphous ice made by condensing the vapour or quenching the liquid.

It seems clear that a new amorphous phase of ice of density about 1.31 g cm⁻³ is produced by the transformation of ice Ih at 77 K and 10 kbar, near its extrapolated melting point. Its volume appears to increase reversibly on removing the pressure and its density decreases to 1.17 g cm⁻³, which is about 26% denser than the films made by condensing the vapour in the range 82–110 K¹¹. On heating, it transforms irreversibly in stages towards a phase that in some ways resembles amorphous ice made by condensing the vapour or quenching the liquid but is clearly different from it, and a discontinuous exothermic change occurs at ~ 117 K at our heating rate of ~ 2.6 K min⁻¹. A wide range of amorphous phases could probably be made by carefully controlled heating.

Amorphous solids can now be made in several ways, and at least four of them have been used to make amorphous phases of ice: condensing the vapour at low temperature, quenching the liquid, transforming the crystal at high pressure below the glass transition of the liquid, and warming the amorphous phase so produced. Each of them can be used to make phases with a range of properties depending on the conditions, and further

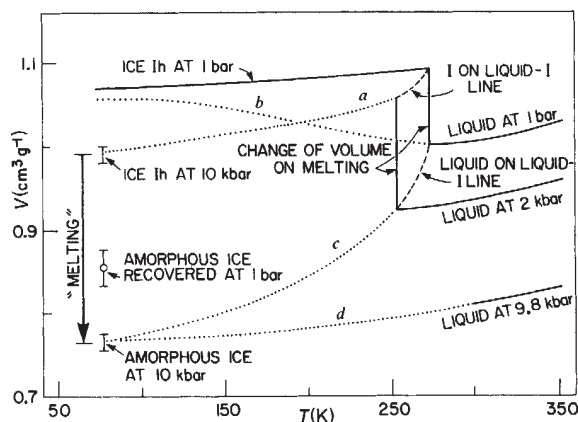


Fig. 3 Graph of the specific volume of ice I at 1 bar, ice I on the liquid-I line (labelled *a*), the liquid at 1 bar (labelled *b*), the liquid along the melting curve from 0 to 2 kbar (labelled *c*), the liquid at 9.8 kbar (labelled *d*), and reasonable extrapolations to 77 K. Measured values are represented by full and dashed lines and extrapolated values by dotted lines. The densities of ice Ih and of the new phase at 77 K and 10 kbar as determined from the compression experiments, and of the new phase (two samples) at 77 K and zero pressure as measured by weighing in liquid and gaseous nitrogen, are also plotted.

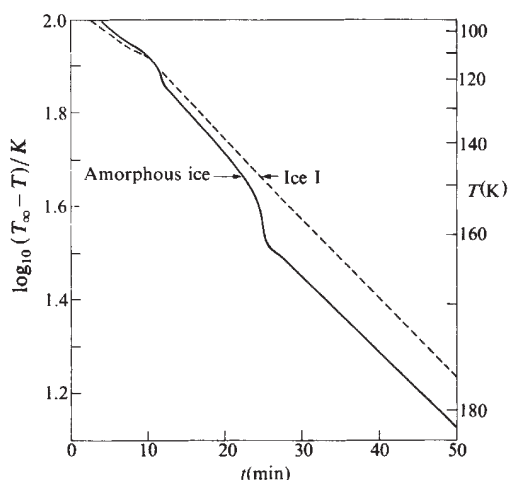


Fig. 4 The first-order plot of a heating curve of the recovered phase. The sample was held in a small silvered glass vacuum flask⁵ immersed in a bath of ethanol and carbon dioxide. The heating rate was $\sim 2.6 \text{ K min}^{-1}$.

studies of them should tell much about how water molecules interact with one another. A possible nomenclature to distinguish the different methods of preparation is amorph-v, amorph-l and amorph-c for the phases made from the vapour, liquid and crystal, respectively, and amorph-c-h for the phases made by heating the phase obtained by transforming the crystal at low temperature.

The transformation crystal-to-amorph-c is obtained surprisingly easily for such a low temperature. This suggests that the crystal becomes unstable and transforms, presumably at its surface, to a phase resembling the supercooled liquid. It is not reversible on the ordinary laboratory time scale and so is not a true melting, and may be viewed as a new kind of transition—an easy transformation of a crystalline solid to a dense amorphous solid.

Similar transformations will probably occur in all solids that have negative volumes of melting when a transformation to a crystalline solid is too slow. Obvious examples are the structure II clathrate hydrates, specifically tetrahydrofuran clathrate hydrate¹², ammonium fluoride I¹³, ammonium fluoride monohy-

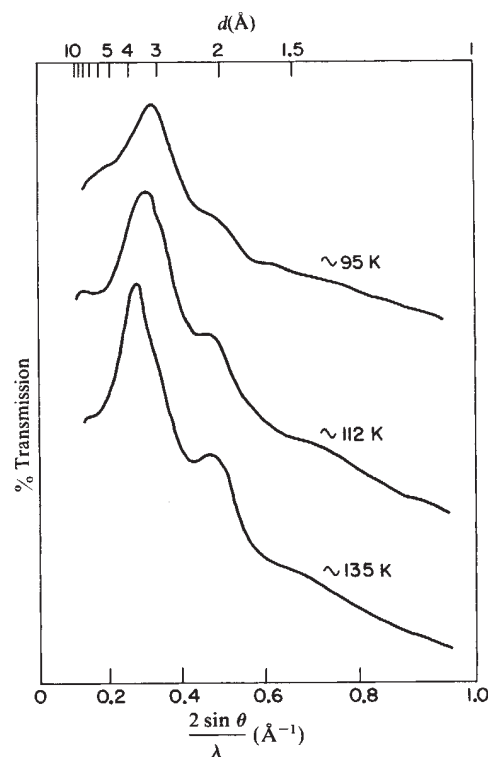


Fig. 5 Representative microphotometer tracings of diffraction patterns of a homogeneous sample of the new phase taken in a flat-plate X-ray camera. The temperature was controlled to $\pm 5 \text{ K}$ by flowing nitrogen gas and was measured by a thermocouple placed near the specimen. Photographs were taken at $\sim 95 \text{ K}$ using zirconium-filtered molybdenum radiation⁵.

drate, indium antimonide¹⁴ and germanium¹⁵, which may transform to an amorphous phase at ~ 10 , ~ 20 , ~ 20 , ~ 50 and $\sim 170 \text{ kbar}$, respectively, at 77 K. Other obvious candidates for amorphization in this way are graphite, diamond, silicon, germanium, cubic and hexagonal boron nitride and phosphide, and graphite-like boron nitride. A transformation to an amorphous phase, rather than to a supposed metallic crystalline form, may limit the maximum force that can be applied to diamond and other tetrahedral compounds at low temperatures and so limit the pressures attainable in diamond anvil apparatuses. The transformation crystal-to-amorph-c in ice and clathrate hydrates may occur in planets formed by the agglomeration of cold particles of ice I or clathrate hydrate.

An obvious way to transform an asymmetrical to a symmetrical hydrogen bond is to squeeze ice I¹⁶. To compress the O-O distance of $2.75 \text{ \AA}$ to the $\sim 2.4 \text{ \AA}$ needed for centrosymmetrical bonds would require only some tens of kilobars. Unfortunately, ice transforms to the amorphous phase at much lower pressures.

National Research Council contribution no. 23539.

Received 20 February; accepted 9 May 1984.

1. Morey, A. W. *The Properties of Glass* 2nd edn, Ch. 1 (Reinhold, New York, 1954).
2. Tammann, G. & Starinkewitsch, J. *Z. phys. Chem.* **85**, 573–578 (1913).
3. Secrist, D. R. & McKenzie, J. D. in *Modern Aspects of the Vitreous State* Vol. 3, 149–165 (Butterworths, London, 1964).
4. De Carli, P. S. & Jamieson, J. C. *J. chem. Phys.* **31**, 1675–1676 (1959).
5. Bertie, J. E., Calvert, L. D. & Whalley, E. *J. chem. Phys.* **38**, 840–846 (1963).
6. Burton, E. F. & Oliver, W. F. *Proc. R. Soc. A* **153**, 166–172 (1938).
7. Dowell, L. G. & Rinfret, A. P. *Nature* **188**, 1144–1148 (1960).
8. Bondot, P. *C. r. heb. Seanc. Acad. Sci., Paris* **265**, 316–318 (1967).
9. Narten, A. H., Venkatesh, C. G. & Rice, S. A. *J. chem. Phys.* **64**, 1106–1121 (1976).
10. Mayer, E. & Brügeller, P. *Nature* **298**, 715–718 (1982).
11. Ghormley, J. A. & Hochandel, C. J. *Science* **171**, 62–64 (1971).
12. Gough, R. & Davidson, D. W. *Can. J. Chem.* **49**, 2691–2699 (1971).
13. Kuriakose, A. K. & Whalley, E. *J. Chem. Phys.* **48**, 2025–2031 (1968).
14. Merrill, L. *J. phys. Chem. Ref. Data* **6**, 1205–1252 (1977).
15. Cannon, J. F. *J. phys. Chem. Ref. Data* **3**, 798–824 (1974).
16. Stillinger, F. H. & Schweitzer, K. S. *J. phys. Chem.* **87**, 4281–4288 (1983).
17. Bridgman, P. W. *Proc. Am. Acad. Arts Sci.* **76**, 9–24 (1945).

A re-interpretation of hydration forces near charged surfaces

Jan J. Spitzer

Department of Chemistry, University of Lethbridge,
Lethbridge, Alberta, Canada T1K 3M4

The concept of hydration forces (bound water) has long been applied to colloid science and its biological applications¹⁻¹¹. While these repulsive forces have been demonstrated experimentally in many cases^{1,2,6-11} our understanding of them remains uncertain^{3-5,12}. Their best definition is perhaps as the discrepancy between experimental results and the currently accepted Gouy-Chapman-Stern/Debye-Hückel-Landau-Verwey-Overbeek (GCS/DLVO) theories. We have found that a new repulsive force is predicted by application of the Debye-Hückel-Manning (DHM) approach to double-layer overlap¹³⁻¹⁵ using the ion-binding/ion-bouncing model (IBBM)¹⁵. This new force arises when overlapping double layers approach each other in such a way that the co-ions are completely expelled and counter-ions are squeezed into a smaller volume. The magnitude of these repulsions is of the order of hundreds of atmospheres at close separations. We therefore question the current interpretations of these forces with the warning that the electrostatic model outlined below needs further elaboration for neutral and weakly charged surfaces^{16,17}.

The DHM approach is based on two assumptions. First, that critical conditions of surface charge density exist which determine the onset of counter-ion adsorption. This assumption has been successfully exploited by Manning in his formulation of limiting laws for polyelectrolytes^{14,18}. Second, the Poisson equation must be used self-consistently, that is, the linearization of the non-linear Poisson-Boltzmann equation is desirable, in principle, to satisfy the linear additivity of electrostatic potentials. The first assumption may be viewed as a consequence of the second requirement.

The DHM approach leads directly to the IBBM model that is shown in Fig. 1 for two overlapping double layers. The solvent molecules are not shown, but it is assumed that their molecular motion averaged over time and space gives rise to a phenomenological dielectric profile. The ions, although shown discretely, are also in molecular motion, and their distribution can, on average, be treated in terms of continuous surface and volume charge densities. We thus have a model to which macroscopic theories of matter (electrostatics, classical thermodynamics, hydrodynamics) can be applied in a self-consistent manner. The main feature of the IBBM model is that adsorption phenomena are necessary to satisfy the electrostatic self-consistency demanded by the theory. The adsorption is assumed to be both two-dimensional, in the plane a (counter-ion monolayer binding), and three-dimensional, in the volume between the planes a and b (counter-ion atmospheric binding). Although this picture is intuitively appealing, it also arises out of necessity for electrostatic self-consistency¹⁵. The electrostatic problem in Fig. 1 is immediately solvable for two limiting cases: (1) $a = b$, no ion-bouncing; and (2) $\sigma_a(+)=0$, no monolayer adsorption, where σ is the primary surface charge density. In general, however, one additional condition must be specified; the atmospheric-monolayer binding condition has worked well in three other unrelated phenomena¹⁵, and is used here also. The condition defines a constant K_a as

$$K_a = \frac{\int_a^b \rho_2 dx}{\sigma_a(+)} \quad (1)$$

where ρ_2 is the counter-ion volume charge density between a and b as given by the DH theory. Figure 1 and one extra-electrostatic condition, such as equation (1), completely define the IBBM model.

When we push the plates in Fig. 1 together at constant primary surface charge density $\sigma_0(-)$, we arrive at a critical separation

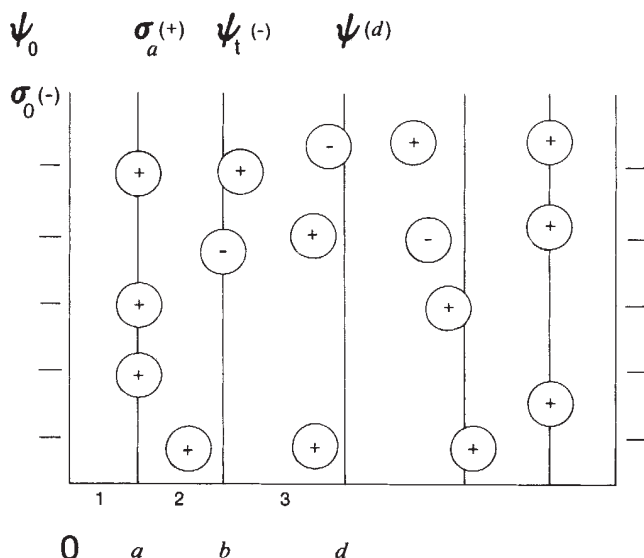


Fig. 1 The overlap of two double layers with negative charge densities $\sigma_0(-)$ that are separated by distance $2d$. The counter-ions are adsorbed as a monolayer $\sigma_a(+)$ at their distance of closest approach a , and also as a volume charge density between the distances a and b . The unknown exclusion boundary b is determined by the thermal electrostatic potential of co-ions $\psi_t^{(-)} = kT/z_e$. At the critical separation distance $d^* = b$ all co-ions are expelled, when there is an onset of a strong electrostatic repulsion as d approaches a . For the data in Table 1, $d^* = 377 \text{ \AA}$.

distance, $d^* = b$. At this point all the co-ions are expelled from between the plates. The d^* is given by

$$d^* = a + \frac{1}{\lambda} \sinh^{-1} \left\{ \frac{1}{\psi_t^{(-)} - \psi_t^{(+)}} \frac{K_a}{1 + K_a} \frac{\sigma_0(-)}{\epsilon_0 \epsilon_2 \lambda} \right\} \quad (2)$$

where ϵ_2 is the local dielectric constant between a and b , and λ (analogous to the DH κ) has been defined elsewhere¹⁹; $\psi_t^{(-)}$ and $\psi_t^{(+)}$ are the thermal electrostatic potentials of the co-ions and the counter-ions¹⁹. When the plates are far apart, $d \gg d^*$, it is easy to recover the GCS/DLVO exponential force law for the repulsions between the plates. However, when the plates are close to each other ($d < d^*$), the midpoint potential $\psi(d)$ is given by

$$\psi(d) = \psi_t^{(+)} + \frac{K_a}{1 + K_a} \frac{\sigma_0(-)}{\epsilon_0 \epsilon_2 \lambda} \frac{1}{\sinh[\lambda(d-a)]} \quad (3)$$

It now follows that the osmotic pressure that pushes the plates apart is

$$\Delta p = \frac{1}{2} \epsilon_0 \psi_t^2 \{ \epsilon_3 \kappa^2 - \epsilon_2 \lambda^2 \} + \frac{1}{2} \epsilon_0 \epsilon_2 \lambda^2 \psi^2(d) \quad (4)$$

This pressure may reach the order of kbars if we try to push the plates together completely ($d \approx a$). The details of these calculations will be given elsewhere but it can be appreciated that in some cases the hydration forces may be coulombic repulsive forces between 'adsorbed' counter-ions when the plates are pushed together in the absence of co-ions ($d < d^*$).

An example of a successful quantitative application of equations (3) and (4) is given in Table 1 using experimental data for montmorillonites for comparison⁹. The quantitative agreement is satisfactory, and the attractive van der Waals' pressure is negligible over the whole range of separation distances (10–80 Å). Similar agreement is to be expected for other clay systems^{6,9}. The sudden increase of the repulsive pressure from 7 to 881 bar on going from 30 to 10 Å separation is notable. These preliminary results suggest that the hydration forces are not separable from the coulombic double layer forces.

To quantify the hydration forces as the difference between the GC/DLVO theory and the experiment is to place too much

Table 1 Comparison between experimental and calculated double-layer pressure

$2d$ (Å)	P_{exp} (bar)	P_{calc}^* (bar)	$P_{\text{calc}}^{\text{w}}$ (bar)
10	—	881	8.95
15	—	71.5	2.14
20	—	24.2	0.74
30	6.7	7.1	0.15
35	4.8	4.7	0.08
40	3.4	3.3	0.05
50	2.0	1.9	0.02
60	1.4	1.3	0.01
70	1.0	0.9	0.00
80	0.7	0.6	0.00

The experimental values are estimates from Fig. 10 in ref. 9 as midpoints between the two bounding curves for a variety of montmorillonites. The conditions are: $\sigma_0 = 14.0 \mu\text{C cm}^{-2}$, $1/\kappa = 305 \text{ Å}$, $1/\lambda = 430 \text{ Å}$. The parameters are: $a = 4.0 \text{ Å}$ (radius of Na^+ ion plus diameter of H_2O molecule), $\epsilon_2 = 78.5$ (dielectric constant of pure water), and $K_a = 0.00585$. The attractive van der Waals' pressures were calculated as in ref. 9.

* At higher concentrations the double-layer repulsion becomes comparable (and smaller) than the van der Waals' attraction.

faith in the nonlinear Poisson–Boltzmann equation (NPBE) on which the GCS/DLVO theory is based. It is well known that the GCS theory is difficult to test experimentally²⁰, and moreover, the NPBE violates the important electrostatic theorem of additivity of electrostatic potentials. Indeed, the following pattern is now emerging from a variety of fields of colloid science (differential capacities²⁰, monolayer studies^{21,22}, double-layer repulsion studies): whenever the GCS theory should show its strength over the linearized Poisson–Boltzmann equation (LPBE), which is also called the Debye–Hückel approximation, some kind of adsorption or postulation of new *ad hoc* effects must be invoked to account for the experiment, particularly at short distances and at high surface charge densities. On the other hand, the IBBM, which is based on the electrostatically self-consistent LPBE, in fact requires that some kind of ionic adsorption must occur at certain critical conditions and beyond. The phenomenological nature of the model (smoothed, mobile charge densities) is also consistent with the models that account for the peculiar dielectric properties of colloid systems^{23–27}. These models are based on the idea of 'surface' conductance, that is, on the mobility of 'bound' ions. Similar ideas are also prevalent in the field of polyelectrolytes¹⁸.

Note that Langmuir derived a special limiting case of a large double-layer force on the assumption of infinitely large surface charge density (absence of co-ions)²⁸. However, the Langmuir limiting force has the incorrect concentration dependence.

I thank Professors B. A. Pethica, J. O'M. Bockris, C. Tanford, J. Lyklema, G. S. Manning, P. F. Low, J. N. Israelachvili, N. H. Sagert, L. G. Hepler, V. A. Parsegian, R. H. Wood, H. P. Bernetto, S. F. O'Shea, K. Ali, D. Dolman and members of this department for critical comments, and Professor J. N. Israelachvili for sending me a preprint. The financial assistance of the Natural Sciences and Engineering Research Council of Canada is acknowledged.

Received 5 January; accepted 23 May 1984.

- Kruyt, H. R. & de Jong, H. G. B. *Kolloidbeihfte.* **28**, 1 (1929).
- Clunie, J. S., Goodman, J. F. & Symons, P. C. *Nature* **216**, 1203 (1967).
- van Olphen, H. *An Introduction to Clay Colloid Chemistry* Ch. 5 (Interscience, New York, 1963).
- Vold, R. D. & Vold, M. J. *Colloid and Interface Chemistry* Ch. 7. (Addison-Wesley, Reading, 1983).
- Franks, F. *Polywater* Ch. 2 (MIT Press, Cambridge, 1981).
- Israelachvili, J. M. & Adams, G. E. *JCS Faraday Trans. 1* **74**, 975–1001 (1978).
- Pashley, R. M. *J. Colloid Interface Sci.* **83**, 531–546 (1981).
- Israelachvili, J. N. & Pashley, R. M. *Nature* **306**, 249–250 (1983).
- Viani, B. E., Low, P. F. & Roth, C. B. *J. Colloid Interface Sci.* **96**, 229–244 (1983).
- Peschel, G., Belouschek, M. M., Müller, M. R. & König, R. *Colloid Polym. Sci.* **260**, 444 (1982).
- Barclay, L., Harrington, A. & Ottewill, R. H. *Kolloidzeitschrift Polym.* **250**, 655–666 (1972).
- Ninham, B. W. *J. phys. Chem.* **84**, 1423–1430 (1980).
- Debye, P. & Hückel, E. *Phys. Z.* **24**, 185 (1923).

- Manning, G. J. *Chem. Phys.* **51**, 924–933 (1969).
- Spitzer, J. J. *Colloid Surfaces* (in the press).
- Le Neveu, D. M., Rand, R. P. & Parsegian, V. A. *Nature* **259**, 601–603 (1976).
- Lis, L. J., McAllister, M., Fuller, N., Rand, R. P. & Parsegian, V. A. *Biophys. J.* **37**, 657–666 (1982).
- Manning, G. *Accts chem. Res.* **12**, 443–449 (1979).
- Spitzer, J. J. *J. Colloid Interface Sci.* **92**, 198–203 (1983).
- Reeves, R. in *Comprehensive Treatise of Electrochemistry* Vol. 1 (eds Bockris, J. O'M., Conway, B. E. & Yeager, E.) Ch. 3, 83–134 (Plenum, New York, 1980).
- Mingins, J., Owens, N. F., Taylor, J. A. G., Brooks, J. H. & Pethica, B. A. *Monolayers Ch.* **3**, 14–27 (Adv. Chem. Ser. No. 144, American Chemical Society, 1975).
- Mingins, J., Taylor, J. A. G., Owens, N. F. & Brooks, J. H. *Monolayers Ch.* **4**, 28–43 (Adv. Chem. Ser. No. 144, American Chemical Society, 1975).
- O'Konski, C. T. *J. phys. Chem.* **64**, 605–609 (1960).
- Chou, S. I. & Shah, D. O. *J. phys. Chem.* **85**, 1480–1485 (1981).
- Schwarz, G. J. *J. phys. Chem.* **66**, 2636–2642 (1962).
- Mandel, M. & Jenard, A. *JCS Faraday Trans. 1* **59**, 2158–2177 (1963).
- Beard, R. B., McMaster, T. F. & Takashima, S. *J. Colloid Interface Sci.* **48**, 92–99 (1974).
- Langmuir, I. *Science* **88**, 430–432 (1938).

Relativistic ^{238}U ion tracks in olivine and cosmic-ray track studies

Claude Perron

Laboratoire de Minéralogie des Roches Profondes et des Météorites, CNRS and Musée d'Histoire Naturelle, 75005 Paris, France

An energetic ion slowing down in an insulating crystal produces radiation damage which can be preferentially etched, making what is called a track¹. Such nuclear tracks have numerous applications in many disciplines, including cosmochemistry and astrophysics¹. Fleischer *et al.*², for example, identified cosmic ray nuclei heavier than Fe by means of fossil tracks recorded in meteoritic crystals. Attempts to use tracks in meteorites to study the chemical composition of ultraheavy cosmic rays, however, have led to ambiguous results (see ref. 3). One possible reason for this lack of success has been the absence of calibrations for ions heavier than krypton, because no particle accelerator could accelerate them to sufficiently high energies. This barrier has been removed by the recent acceleration, at the Bevalac of the Lawrence Berkeley Laboratory, of uranium ions to relativistic energies^{4–6}. The results on tracks of ^{238}U ions of 190 MeV per nucleon in olivine that we now report, are the first of this kind obtained with a mineral nuclear track detector. They show that the etchable part of a U ion trail is surprisingly long (nearly 3 mm). These results have an important bearing both on the interpretation of cosmic-ray tracks in lunar and meteoritic crystals, and on models of track formation in minerals.

Only part of the ion path can give a track, that where the radiation damage is the highest, near the end of the range. The heavier the ion, the longer the part of the ion path that can be revealed. Thus, in principle, the measurement of the length of a track should permit the identification of the atomic number, Z , of the ion which induced it, if the track is contained entirely within the volume of the crystal (that is, it does not intersect the surface). Cracks or cleavages are generally used to allow the etchant to reach such a track, as in the TINCLE technique of Lal *et al.*⁷. Obviously, it is important to measure the track lengths corresponding to known ions. Until recently, this was impossible for all ions heavier than Kr ($Z = 36$). The Bevalac transformation^{4–6} has changed the situation, and allowed us to observe tracks produced in olivine by U ions of relativistic energy.

Olivine ($\{\text{Mg, Fe}\}_2\text{SiO}_4$) was chosen because large crystals of this mineral are found in some meteorites which can be used for ultraheavy cosmic-ray track studies³. Large olivine crystals (several millimetres in size) of both meteoritic and terrestrial origin were mounted in epoxy resin and polished, then irradiated, at the Bevalac, by U ions under different geometries. The U ions, accelerated to 260 MeV per nucleon, had to traverse a 60 g cm^{-2} polyethylene beam pipe window and 1.2 m of air, before reaching the targets. In addition, a 0.5- (or 0.87-) mm thick lead absorber was placed in front of the targets to degrade

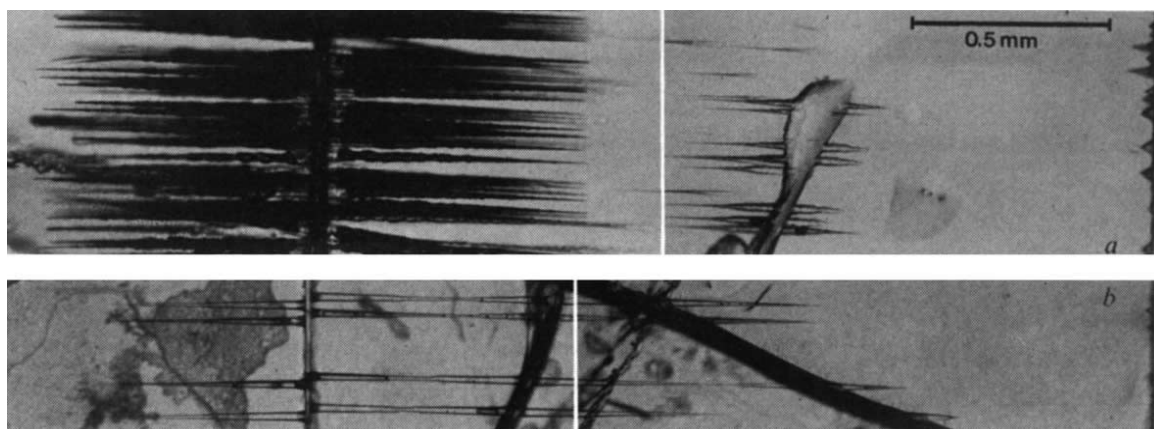


Fig. 1 Photomicrographs of 190 MeV per nucleon U ion tracks in two olivine crystals, etched for 167 h (a) and 127 h (b). The U ions entered the crystals from the right edge, where etch pits can be seen. The crystals were sawn before irradiation to permit revelation of the tracks. In b, the two pieces of crystal have been slightly misaligned after etching.

the ion energy and reduce the range of the ions to the dimensions of the crystals. The energy at entry of the olivine crystals is calculated to be 190 (or 155) MeV per nucleon (ref. 8). In type 1 targets, crystals were irradiated perpendicularly to the polished surface. As the energy of the incident ions was expected to be too high for the tracks to be etchable from the ion-bombarded surface, the crystals were cut, before the irradiation, at about 2 mm from that surface, so that the etchant could reach the etchable part of the tracks, in a way similar to the TINCLE technique⁷. The two new surfaces were polished, then put back into close contact with each other for the irradiation. After etching both crystal pieces, the tracks could be observed with a microscope through the top or the bottom surfaces (also polished). On targets of type 2, the U beam made an angle of 35° with the polished surface of the crystals. After irradiation, the crystals were repolished, the new surface making a small angle (~17°) with the previous one. In this way, the crystal surface intersects the U ion trails at all values of the residual range, from 0 to the range of the incoming ions. This method^{9,10} permits one to measure the track etch rate as a function of the residual range. The crystals were etched in the etchant of Krishnaswami *et al.*¹¹, modified as in ref. 3, and the pH was adjusted to give the largest values of both the track etch rate and the ratio of track etch rate to bulk etch rate (M. Maury and C. Perron, in preparation).

Figure 1 shows a photomicrograph of two crystals of type 1, with very different track densities, etched for 167 h (a) and 127 h (b), after irradiation by 190 MeV per nucleon U ions. As expected, tracks are seen on both sides of the cut. In Fig. 1a, the total etched length is ~1,350 μm , but short etch pits are visible on the edge where the ions penetrated the crystal. This means that the tracks are, in fact, etchable all the way from the entrance surface to the end of the range, that is, to a length of 2,770 μm , or a few tens of micrometres more, if one takes into account the thickness of crystal which has been etched away. This is confirmed by the presence of shorter tracks, revealed through a natural crack in the crystal, or from the upper surface, and aligned with longer portions, etched from the cut. In Fig. 1b, several cracks allowed some tracks to be revealed on the major part of their etchable length. In fact, the total etchable length of U tracks must be longer than what we measure, since the tracks are already visible at the surface. However, the etch pits are only 20–70 μm long after a 167-h etching, indicating an etch rate so low that we must be very close to the actual end of the etchable portion of the tracks.

In principle, it should be possible to reveal the tracks completely by simply etching them for a longer time. In practice, this is not so simple. Figure 2 shows the results of track length measurements performed on two fragments of a crystal of type 2, irradiated by 155 MeV per nucleon U ions. The two sets of

points represent the variation of the etched track length as a function of residual range R , after etching for 3 h and 16 h. After the shorter etch, no tracks were revealed for $R > 1,200 \mu\text{m}$. It is on the basis of such results that it has been erroneously argued that the total etchable length of U tracks was ~1,300 μm (ref. 12, 13). Below $R \sim 400 \mu\text{m}$, the etched track length (and thus the etch rate) is roughly constant, a feature predicted by the track model of Dartyge *et al.*¹⁴. After the longer etch, tracks are revealed at all R , up to 2,000 μm (that is, the range of 155 MeV per nucleon U ions in olivine), but now the etch-rate plateau extends up to 1,000–1,200 μm .

Three conclusions can be drawn from these observations: (1) there is a considerable increase in etch rate, for $R > 1,000 \mu\text{m}$, after the first 3 h of etching, or, there is a time delay before etching really starts, which is longer at high residual range; (2) conversely, there is a considerable decrease of etch rate as the tracks get longer: for $R < 400 \mu\text{m}$, the part of the tracks etched during the first 3 h has about the same length as the additional part etched during the 13 h following; (3) as a consequence of this last point, the etch rate becomes more and more dependent on the track length that the etchant has to pass through, and less and less dependent on the residual range; hence the etch rate plateau. Part of the plateau observed after the 3-h etch is due to this effect—although a plateau due to a saturation of the radiation damage does exist, as evidenced by experiments made with low-energy U ions—since the etch rate (~25 $\mu\text{m h}^{-1}$) is already significantly lower than the value observed for a very short etch (~60 $\mu\text{m h}^{-1}$ for a 20 mn etch). The last two points are quite understandable: it obviously gets more and more difficult to renew the etchant through the long and narrow tunnels that go to make up the tracks, so that the etch rate decreases with time, and finally tends to zero, whatever the radiation damage. Although trivial, this is perhaps the most important point of this study. It is, by itself, sufficient to explain why no reliable quantitative data on ultraheavy cosmic rays have ever been obtained from the study of tracks in meteorites: the measured track length distributions reflect essentially the difficulty of etching long tracks, which varies with track orientation, and most probably the chemical composition of the crystals, and very little the charge composition of the ions which produced the tracks.

Is this method then doomed to failure? The answer is unclear, because other factors also have a role^{12,13}, but in any case, one has to realize that the etchable part of the tracks must be fully revealed, if one wants to identify the ions which induced them. One way of doing this is to take advantage of cracks in the crystal (Fig. 1b). The problem is that it is not usual to find crystals with many regularly-spaced natural cracks. A solution might be to use heavy-ion irradiations through very narrow slits, at ~90° to the long tracks one wants to reveal, and at regular

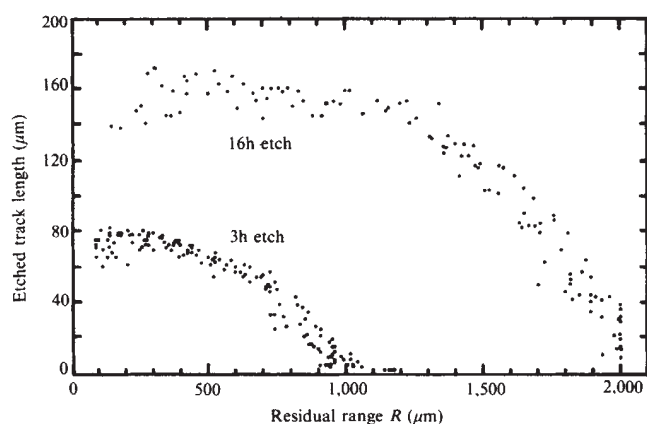


Fig. 2 Measured etched track lengths as a function of residual range R , for 155 MeV per nucleon U ions, in two pieces of the same olivine crystal, etched for 3 h (lower points) and 16 h (upper points). Points corresponding to fully etched tracks (at low values of R) have been omitted for clarity.

intervals, all over the crystal surface. Very encouraging preliminary results have been obtained, using 14 MeV per nucleon Xe ions from the GSI (Darmstadt, FRG) accelerator. The need for such an etching from several points along the ion paths is supported by the variations in 'diameter' of some tracks (Fig. 1b), which correspond to places where the longitudinal etching has been blocked for some time, while the etched channels were enlarging. This etch-blocking phenomenon is not clearly understood, but it is not unusual in olivine¹⁰.

Models based on the existence of a primary ionization threshold for track registration predict a length of $\sim 800 \mu\text{m}$ for U ion tracks in meteoritic minerals¹, in striking disagreement with our observation. This confirms that the idea of an ionization threshold is wrong^{10,14}. However, the length of $\sim 800 \mu\text{m}$ is deduced from a threshold value based on data on light ion tracks ($Z \leq 26$). Note that krypton ($Z=36$), which produces tracks in olivine below 10–11 MeV per nucleon (ref. 10), at this energy induces about the same primary ionization as a 180–200 MeV per nucleon U ion. The threshold notion may thus be a useful approximation, provided it is applied to a limited charge range. For Pb ions, one then predicts a total etchable length of 1,600–1,800 μm . This should be experimentally verified. If true, it would mean that the track method potentially has an excellent charge resolution to separate the cosmic ray U-group from the Pb-group, in meteoritic crystals.

I thank H. J. Crawford and D. E. Greiner for help with the Bevalac irradiations, C. B. Moore, V. P. Pereygin and H. J. Schubnel for the olivine samples, M. Bourot-Denise for help with the preparation of the targets, and P. Pellas for enthusiastic and fruitful support. Critical comments by S. A. Durrani are much appreciated.

A seismic Q -profile for the lower mantle from short-period P waves

Michael J. Shore*

Department of Earth and Planetary Sciences, The Johns Hopkins University, Baltimore, Maryland 21218, USA

A few studies of the quality factor Q (the ratio of peak-to-dissipated energies of seismic waves) in the Earth's mantle have used a limited data set of spectral ratios for body waves^{1–3}. Most estimates of Q have come from the analysis of surface waves and free oscillations, but use of such data to estimate Q in the deep mantle suffers from a significant lack of resolution. To improve this resolution in the lower mantle, we have compiled a large data base of digital, short-period P-wave recordings from underground nuclear explosions. We have analysed P and PP waves to develop a set of t^* (seismic travel time divided by effective path Q) station corrections corresponding to the upper mantle beneath each station and a t^* -distance profile. The t^* -distance profile is inverted to determine a Q -depth profile for the lower mantle. A thermal activation model used to interpret the Q -depth profile and to estimate a lower mantle geotherm, enables a thermal boundary layer to be identified in the lowermost 200 km of the mantle. This boundary layer is large enough to be indicative of mantle convection driven by heat flow from the Earth's core.

Currently, t^* is in widespread use to describe the attenuation of body waves in the Earth. In the frequency domain, the attenuation operator acts on the seismic signals in the form:

$$S(f) = S_0(f) e^{-\pi f t^*} \quad (1)$$

where S_0 is the unattenuated signal, S is the attenuated signal and f is frequency. Analysis of two P waves from a common event permits the determination of the difference in t^* (dt^*) for the two paths, assuming Q is independent of frequency in the 0.5–5.0 Hz passband. This is possible because the observed signal amplitude spectrum (S) is composed of the source function (C) modified by t^* in the upper mantle beneath the source (U), a distance-dependent t^* (D) caused by attenuation in the middle and lower mantle, t^* in the upper mantle beneath the recording site (R), and geometrical spreading (G) according to

$$S(f) = GC(f) e^{-\pi f (U+D+R)} \quad (2)$$

Ignoring geometrical spreading which does not change the shape of the spectrum, the ratio (A) of the spectra of the two P waves from a common event is

$$A(f) = e^{-\pi f (D_1+R_1-D_2-R_2)} \quad (3)$$

dt^* for the two P waves is therefore

$$dt^* = D_1 + R_1 - D_2 - R_2 \quad (4)$$

and can be computed from the slope of the log of the spectral ratio in the frequency domain.

Analyses of the spectra of 147 P waves from 31 underground nuclear explosions in the western Soviet Union, recorded at 22 stations in North America, provided 341 measurements of dt^* . These measurements utilized the maximum usable bandwidth and averaged 0.6–3.4 Hz. These data are inverted using the successive approximation technique to determine upper mantle t^* values for each station (Fig. 1) and a distance- t^* profile (Fig. 2). The determination of absolute station corrections is possible by adjusting them so that a reference station on the Canadian Shield is 0.0. The station terms therefore measure the anomalous attenuation under each station relative to what is observed for the high- Q upper mantle under the shield. It is possible to determine an absolute value for the distance- t^* relationship, as recordings of P waves at the bounce points of PP waves were

* Present address: Shock Physics Directorate, Defense Nuclear Agency, Washington DC 20305, USA.

Received 16 March; accepted 4 June 1984.

1. Fleischer, R. L., Price, P. B. & Walker, R. M. *Nuclear Tracks in Solids* (University of California Press, 1975).
2. Fleischer, R. L., Price, P. B., Walker, R. L., Maurette, M. & Morgan, G. *J. geophys. Res.* **72**, 355–366 (1967).
3. Otgonsuren, O. et al. *Astrophys. J.* **210**, 258–266 (1976).
4. Alonso, J. R. et al. *Science* **217**, 1135–1137 (1982).
5. Heckman, H. H., Karant, Y. J. & Friedlander, E. M. *Science* **217**, 1137–1138 (1982).
6. Ahlen, S. P., Tarle, G. & Price, P. B. *Science* **217**, 1139–1140 (1982).
7. Lal, D., Rajan, R. S. & Tamhane, A. S. *Nature* **221**, 33–37 (1969).
8. Barkas, W. H. & Berger, M. J. *Tables of Energy Losses and Ranges of Heavy Charged Particles* (NASA SP-3013, 1964).
9. Storzer, D., Poupeau, G. & Kratschmer, W. *Geochim. cosmochim. Acta Suppl.* **4**, 2363–2377 (1973).
10. Price, P. B., Lal, D., Tamhane, A. S. & Pereygin, V. P. *Earth planet. Sci. Lett.* **19**, 377–395 (1973).
11. Krishnaswami, S., Lal, D., Prabhu, N. & Tamhane, A. S. *Science* **174**, 287–291 (1971).
12. Perron, C. & Pellas, P. *18th int. Cosmic Ray Conf., Bangalore Pap.* OG 1-24 (Tata Institute of Fundamental Research, Bombay, 1983).
13. Pellas, P. & Perron, C. *Nucl. Instrum. Meth. Phys. Res.* **B1**, 387–393 (1984).
14. Dartyge, E., Duraud, J. P., Langevin, Y. & Maurette, M. *Phys. Rev.* **B23**, 5213–5229 (1981).

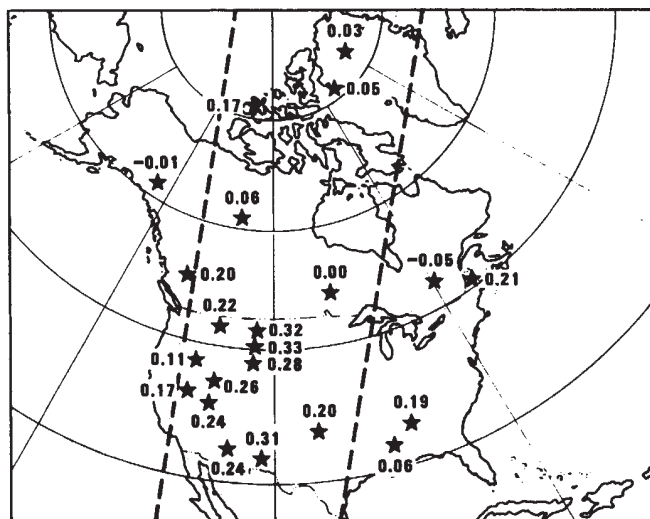


Fig. 1 t^* station corrections. Station corrections in seconds have been adjusted so that the Canadian Shield site is 0.0. The corrections therefore measure the anomalous attenuation of the upper mantle relative to a typical shield site.

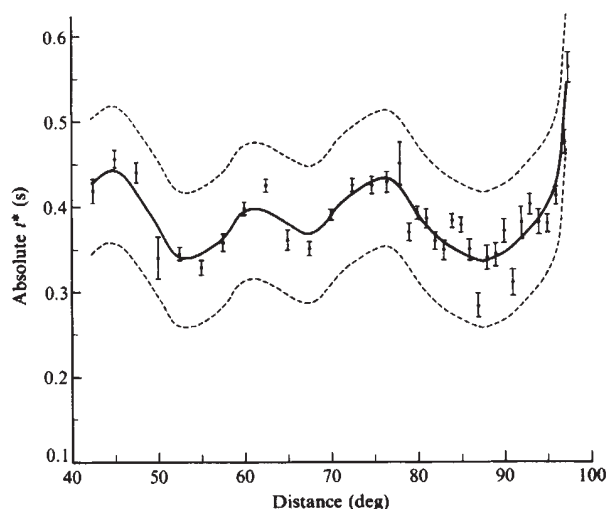


Fig. 2 t^* plotted against distance. Error bars indicate the 95% confidence limits on the shape of the curve. Dashed lines indicate the 95% confidence limits on the absolute level of the curve.

identified in the data base. Assuming that the recorded P wave at the bounce point is representative of the reflected wave, it is possible to measure absolute t^* on the path from the bounce point to a more distant station recording the PP wave, using the spectral ratio of the two signals. The distance- t^* profile is inverted to determine a Q -depth profile (Fig. 3) using a ray tracing formulation⁴ and a specific velocity model⁵. Included with the new Q -depth profile are two earlier determinations. SL8⁶ was determined using free oscillation data and is observed to agree in overall level above depths of 2,000 km, but clearly disagrees below that depth. SL1⁷ is similar to SL8, except that body wave constraints were included in its determination. Model SL8 indicates a higher Q in the lowermost mantle which is more in line with the results of this study. Observed differences in long- and short-period Q estimates lead directly to the current discussions of frequency-dependent Q . All three models indicate a dramatic drop in Q in the lowermost 200 km of the mantle, commonly known as D'' . Diffraction of the waves by the core complicates the interpretation in D'' especially for the long-period studies. Because this study used short-period P waves,

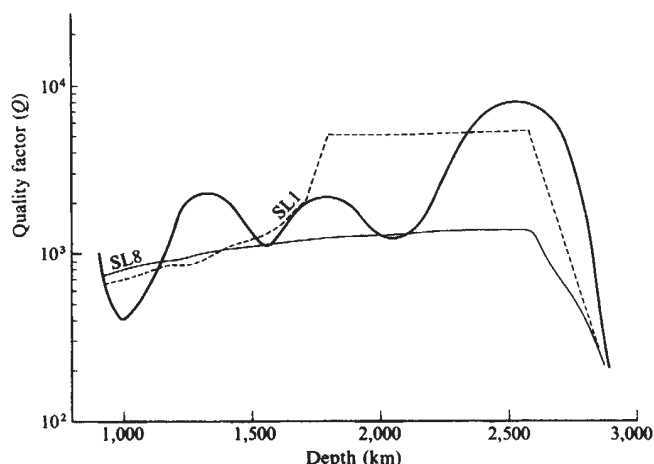


Fig. 3 Comparison of Q models. SL8 model is derived from free oscillation data. SL1 is derived from free oscillation data with some constraints as a result of body wave analysis. The dark curve is obtained from the exclusive analysis of P waves and PP waves.

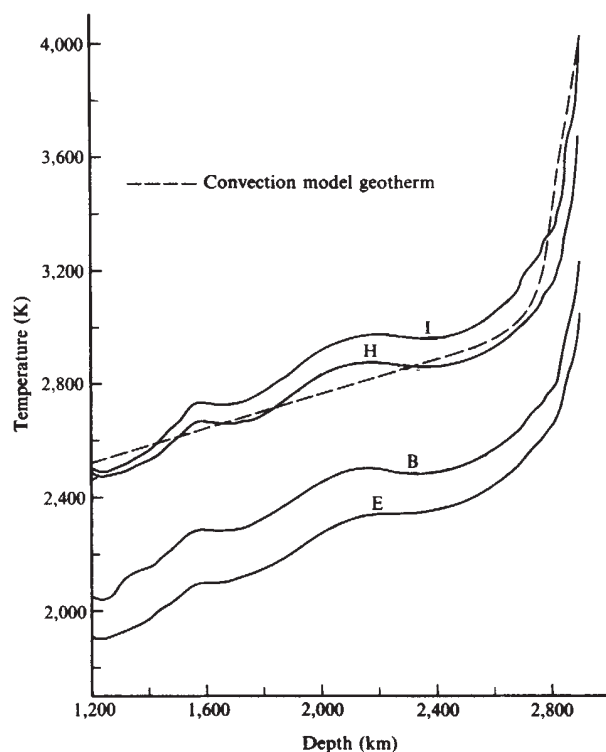


Fig. 4 Lower mantle geotherms for thermal activation model. Curves B and E represent cool mantle models while curves H and I represent warm mantle models. Note the strong thermal boundary layer at the base of the mantle and its general agreement with the convecting mantle model.

diffraction effects contaminate the results only beyond 96° which corresponds to waves bottoming in the lowermost 20 km of D'' . Many investigators⁸⁻¹¹ have examined other aspects of seismic wave propagation in the lowermost mantle and indicate that its fundamental properties are different from most of the mantle. Chemical change and/or thermal boundary layers have been offered as possible interpretations for this change in observed mantle properties.

Independent of the specific mechanism^{12,13} which determines Q in the mantle, a temperature-pressure dependence is anticipated. This dependence is postulated to be a thermal activation

model which is expressed by

$$Q = Q_0 \exp \left(\frac{E^* + PV^*}{RT} \right) \quad (5)$$

where E^* is the activation energy, V^* is the activation volume, T is temperature, P is pressure and Q_0 and R are constants. There is no intention to exclude other models for attenuation. The advantage of the thermal activation model is that it utilizes parameters which are constrained by other geophysical evidence.

In this analysis, the pressure–depth profile¹⁴ is not varied and the temperature above D'' is obtained by extrapolating the 1,000-km temperature of 1,800–2,500 K with an adiabatic gradient of 0.3–0.4° km⁻¹. Using these constraints, a family of models is determined which fit the Q -depth profile and require an activation volume of 2.0–2.5 cm³ mol⁻¹ and an activation energy of 90–115 kcal mol⁻¹. These values are well within the allowable ranges determined from other independent measurements^{13,15–17}. The geotherm (Fig. 4) indicates that most of the lower mantle is adiabatic which is consistent with a convection

model. A rapid increase in temperature is observed near the core–mantle boundary. This is best interpreted as a thermal boundary layer and predicts a core–mantle boundary temperature of 2,900–3,300 K for cool mantle models (curves B and E; 1,800 to 2,000 K at 1,000-km depth). This core–mantle boundary temperature is in agreement with extrapolations of the Fe–FeS eutectic melting temperature¹⁸. The warm mantle models (curves H and I; 2,400 K at 1,000-km depth) predict a core–mantle boundary temperature of 3,500–4,000 K and are in reasonable agreement with published convection models for mantle convection driven by heat flow from the Earth's core¹⁹. The appearance of a single boundary layer at the base of the mantle argues against convection in a chemically distinct D'' , which would possess two boundary layers²⁰, assuming the thermal activation model predominates in determining the observed Q and there is no depth dependence in E^* and/or V^* . Other physical models for Q in the mantle could permit a chemically distinct D'' consistent with the observed Q -depth profile.

I thank Drs Peter Olson, Ralph Alewine and Bruce Marsh for helpful suggestions and assistance.

Received 10 June 1983; accepted 18 May 1984.

1. Teng, T. *J. geophys. Res.* **73**, 2195–2208 (1968).
2. Frazier, C. W. & Filson, J. *J. geophys. Res.* **77**, 3782–3787 (1972).
3. Sipkin, S. A. & Jordan, T. H. *Bull. seism. Soc. Am.* **70**, 1071–1102 (1980).
4. Julian, B. J. & Anderson, D. L. *Bull. seism. Soc. Am.* **58**, 339–366 (1968).
5. Herrin, E. *et al. Bull. seism. Soc. Am.* **58**, 1193–1241 (1968).
6. Anderson, D. L. & Hart, R. S. *J. geophys. Res.* **83**, 5969–5882 (1978).
7. Anderson, D. L. & Hart, R. S. *Phys. Earth planet. Inter.* **16**, 289–306 (1978).
8. Bolt, B. A. *Phys. Earth planet. Inter.* **5**, 301–311 (1972).
9. Cleary, J. L. *Phys. Earth planet. Inter.* **9**, 13–27 (1974).
10. Doornbos, D. J. *Geophys. J. R. astr. Soc.* **44**, 447–470 (1976).
11. Ruff, L. J. & Helmberger, D. V. *Geophys. J. R. astr. Soc.* **68**, 95–119 (1982).
12. Jackson, D. D. & Anderson, D. L. *Rev. Geophys. Space Phys.* **8**, 1–63 (1970).
13. Minster, J. B. in *Proc. Enrico Fermi Int. School Phys.* (eds Dziewonski, A. & Boschi, E.) (Academic, New York, 1980).
14. Dziewonski, A. *et al. Phys. Earth planet. Inter.* **10**, 12–48 (1975).
15. Minster, J. B. & Anderson, D. L. *Phil. Trans. R. Soc.* **299**, 319–356 (1981).
16. Karato, S. *Phys. Earth planet. Inter.* **22**, 155–161 (1980).
17. Kohlstedt, D. L. *et al. J. geophys. Res.* **85**, 3122–3130 (1980).
18. Stacey, F. D. *Phys. Earth planet. Inter.* **15**, 341–348 (1977).
19. Elsasser, W. M. *et al. J. geophys. Res.* **84**, 147–155 (1979).
20. Jeanloz, R. & Richter, F. M. *J. geophys. Res.* **84**, 5497–5504 (1979).

Two distinct trends for cooling of the oceanic lithosphere

Anny Cazenave

Groupe de Recherche de Géodésie Spatiale,
Centre National d'Etudes Spatiales, 18 Avenue Edouard Belin,
31055 Toulouse Cedex, France

Cooling and contraction of the oceanic lithosphere with age is reasonably well described by simple thermal models. Both 'plate' models^{1–4} and 'half-space' models^{5–7} give similar results up to ages of 50–70 Myr but predict quite different behaviour for sea floor depth, heat flow and other parameters in old basins. Tests of thermal models have been mainly based on topography data^{3,8,9} which are biased by anomalous shallow regions. Altimeter-equipped satellites offer a new type of observable quantity—geoid height data—which can also constrain lithospheric cooling models¹⁰. We present here new observational constraints on the thermal cooling models of the oceanic lithosphere, based on analysis of geoid data over the whole Pacific Ocean. We find that the data closely follow the behaviour predicted by plate models and that they suggest the existence of two distinct cooling regimes, one for sea floor ages ≤ 30 Myr, the other for ages > 30 Myr.

Cooling of the lithosphere causes the geoid height to decrease uniformly with increasing plate age, symmetrically away from the ridge crest^{11–13}. This geoid slope is, in turn, responsible for the observed geoid offset across fracture zones because of the resulting abrupt change in plate age on the two sides^{14–17}. While the change in geoid height due to lithospheric cooling is typically 5–10 m over distances of 1,000–2,000 km, it is very difficult to isolate it from other unrelated long wavelength anomalies. The geoid offset at fracture zones has a step-like signature, downward from the younger, shallower side to the older, deeper side, with amplitude ranging from a few tens of a centimetre to a few metres over distances of 100–150 km. This characteristic is mostly easily detectable and presumably has little contamination from features unrelated to cooling. This geoid offset, Δh , divided

by the age offset, Δt , approximates the derivative with respect to age of the geoid slope, $h(t)$, due to lithospheric cooling.

Thus, if the age of each side of the fracture zone is known, the observed dependence of $\Delta h/\Delta t$ on age can be used to compare the theoretical curves derived from cooling models, allowing some model parameters to be estimated.

We have previously¹⁸ analysed geoid height data across the Udintsev fracture zone (south Pacific) derived from the Seasat satellite and found that geoid data were incompatible with half-space models and followed the behaviour predicted by plate models. We derived two parameters: (1) the thermal thickness, H , of the plate; and (2) the product of three quantities: bottom boundary temperature, T_m , thermal diffusivity, κ , and volume coefficient of thermal expansion, α . To fit the data associated with plate ages ≤ 30 Myr, a plate thickness, H , in the range 50–70 km was estimated whereas for larger ages (> 30 Myr) geoid observations were better explained by a larger value of H in the range 70–90 km. We interpreted the small plate thickness found for ages ≤ 30 Myr as evidence of a local thermal anomaly possibly linked to the presence of a hotspot in the Eltanin region. We have extended this approach to many other fracture zones of the south Pacific and north-east Pacific (details of the latter analysis will be presented elsewhere¹⁹). More than 15 different fracture zones have been considered and ~ 85 geoid offsets have been collected over a large area extending from 60° S to 40° N and 80° W to 160° W in the Pacific, covering a plate age range of ~ 60 Myr. In addition to these, four offsets from the Amsterdam fracture zone in the south-east Indian Ocean are also considered here. The quantity $\Delta h/\Delta t$ is estimated using the method proposed by Crough¹⁴, and described in ref. 18, to measure the geoid step: examples of geoid offsets are shown in Fig. 1. Then, using all these $\Delta h/\Delta t$ estimates, the observed dependence of $\Delta h/\Delta t$ on age has been derived. A log–linear plot of this relationship is shown in Fig. 2.

It is easy to show theoretically that $\Delta h/\Delta t \approx 2\pi G/g\rho_m\alpha\kappa T_m$ at all ages in the half-space models and at $t=0$ in plate models while in the latter, $\Delta h/\Delta t \approx 4\pi G/g\rho_m\alpha\kappa T_m \exp(-\pi^2\kappa t/H^2)$ at larger ages (G , gravitational constant; g , mean surface gravity; ρ_m , upper mantle density). Thus, in the latter case, $\ln(\Delta h/\Delta t)$ decreases linearly with age. The theoretical dependence of

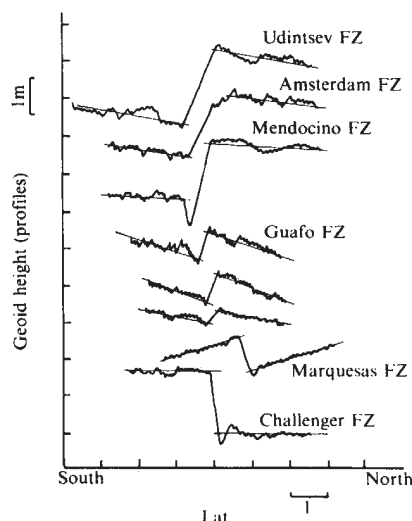


Fig. 1 Profiles of geoid offsets across fracture zones. The step is measured between the two parallel solid lines representing the regional trend.

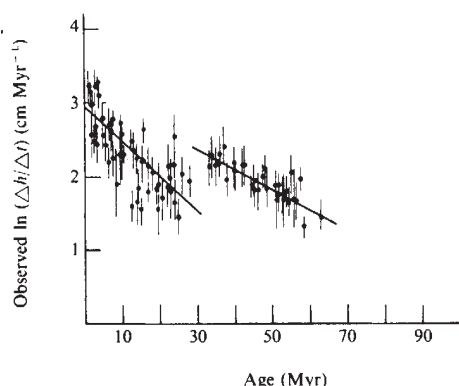


Fig. 2 Observed $\ln(\Delta h/\Delta t)$ plotted against age. The geoid offset Δh has been measured across many fracture zones of the Pacific Ocean. Four data points are associated with the Amsterdam fracture zone (south-east Indian Ocean). Solid lines are the regression lines fitting the data.

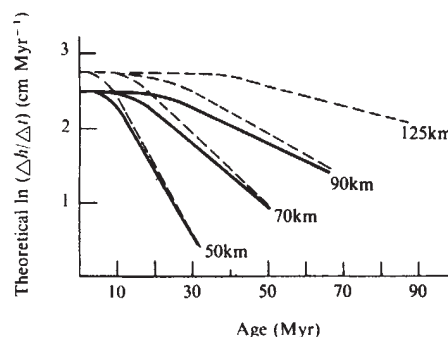


Fig. 3 Theoretical $\ln(\Delta h/\Delta t)$ versus age curves ('plate' model) computed for different values of the thermal thickness, H . Solid curves correspond to the (α, κ, T_m) values fitting the geoid data while dashed curves correspond to the (α, κ, T_m) values determined by Parsons and Sclater³. The dashed curve with $H = 125$ km corresponds to the preferred model of these authors.

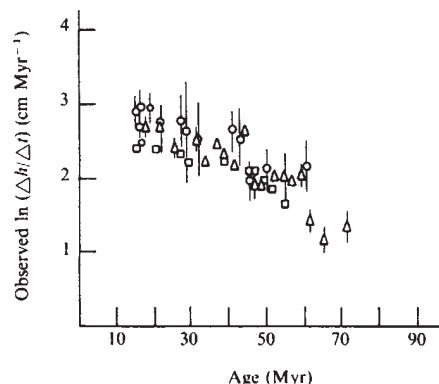


Fig. 4 Plot of observed $\ln(\Delta h/\Delta t)$ versus age using geoid offset data across the Mendocino fracture zone, estimated in earlier analyses¹⁵⁻¹⁷. No error bars were produced for some of these data. Data sources: Δ , ref. 15; $\circ$, ref. 16; $\square$, ref. 17. Note that the Sandwell and Schubert data¹⁶ are systematically higher than the other data, possibly as a result of the method adopted to estimate the geoid step.

$\ln(\Delta h/\Delta t)$ on age computed for different values of the plate thickness, H , and two sets of the parameters (α, κ, T_m) is shown in Fig. 3. These curves refer to plate models because half-space models predict a horizontal straight line for a given set of (α, κ, T_m) .

From Fig. 2, we note that: (1) the data show the linear decrease with age predicted by plate models, as long as t is not too small; (2) there seem to be two distinct linear trends, one characterizing the age range 0–30 Myr, the other, the age range 30–60 Myr. A single set of thermal parameters cannot fit the whole data set; rather, the two distinct trends could be explained by two distinct values of the plate thickness, H , as proposed from geoid data across the Udintsev fracture zone only¹⁸. These data contribute to ~27% of the present data set and if removed, the two trends are still observed.

We fitted regression lines by least squares to the two groups of data (0–30 Myr and 30–60 Myr), as shown by solid lines in Fig. 2. The slopes of the regression lines give the quantity κ/H^2 . Thus, H can be determined if κ is assumed to have less uncertainty than H . The intercept of the regression line gives the value of $\alpha\kappa T_m$, and thus gives αT_m if κ is known. For the two age ranges 0–30 Myr and 30–60 Myr, we deduce plate thickness values of 66 ± 5 km and 92 ± 6 km, respectively. Earlier analyses of geoid offsets across fracture zones were all devoted to the Mendocino fracture zone (north-east Pacific)¹⁴⁻¹⁷. These studies

considered a very limited selection of points (between 11 and 18, compared with the 85 data points shown in Fig. 2, <20% of the data set considered here). Moreover, these analyses had only a few points (3–7) associated with ages <30 Myr, with a minimum age of ~16 Myr. Thus, it is not surprising that these preliminary investigations did not show the two trends reported here. However, they do show a similar linear decrease to our more global data set for plate ages older than 30 Myr (see Fig. 4).

The present results indicate that geoid observations favour plate cooling models (which predict that sea floor depth departs significantly from the linear increase with the square root of age beyond ~70 Myr in approaching a constant asymptotic value). Such a model requires a supply of heat at the base of the slab if a balance is to be reached in old basins between the heat supplied from below and the heat lost by conduction through it. The geoid-derived thermal thickness differs significantly from the thickness ($H = 125$ km) determined by Parsons and Sclater³ using topography data in the north Pacific (the $\ln(\Delta h/\Delta t)$ curve constructed with the Parsons and Sclater parameters is plotted in Fig. 3). The latter parameters are widely used as reference values to define the standard sea floor subsidence with respect to which depth anomalies are estimated. However, the depth-age relationship deduced from the geoid-derived parameters predicts a lesser sea floor subsidence, and thus smaller positive

depth anomalies than commonly reported. For plate ages < 30 Myr, the thermal thickness of the slab appears ~ 25 km thinner than beyond 30 Myr, suggesting that the T_m isotherm is uplifted in the younger portion of the plate. This may be interpreted in terms of thermal perturbation near the ridge crest.

This study was supported by the Centre National d'Etudes Spatiales.

Received 9 January; accepted 1 June 1984.

1. McKenzie, D. P. *J. geophys. Res.* **82**, 6261–6273 (1967).
2. Selater, J. G. & Francheteau, J. *Geophys. J. R. astr. Soc.* **20**, 509–542 (1970).
3. Parsons, B. & Selater, J. G. *J. geophys. Res.* **82**, 803–827 (1977).
4. Bottinga, Y. & Allègre, C. *Tectonophysics* **32**, 9–59 (1976).
5. Turcotte, D. L. & Oxburgh, E. R. *J. Fluid Mech.* **28**, 29–42 (1967).
6. Parker, R. L. & Oldenburg, D. W. *Nature phys. Sci.* **242**, 137–139 (1973).
7. Oldenburg, D. W. *Geophys. J. R. astr. Soc.* **43**, 425–451 (1975).
8. Davis, E. E. & Lister, C. R. B. *Earth planet. Sci. Lett.* **21**, 405–413 (1974).
9. Hayes, D. E. *EOS* **64**, 170 (1983).
10. Lister, C. R. B. *Geophys. J. R. astr. Soc.* **68**, 219–240 (1982).
11. Haxby, W. F. & Turcotte, D. L. *J. geophys. Res.* **83**, 5473–5478 (1978).
12. Parsons, B. & Richter, F. M. *Earth planet. Sci. Lett.* **51**, 445–450 (1980).
13. Sandwell, D. & Schubert, G. *J. geophys. Res.* **85**, 7235–7241 (1980).
14. Crough, S. T. *Earth planet. Sci. Lett.* **44**, 224–230 (1979).
15. Detrick, R. S. *J. geophys. Res.* **86**, 11751–11762 (1981).
16. Sandwell, D. & Schubert, G. *J. geophys. Res.* **87**, 3949–3958 (1982).
17. Cazenave, A., Lago, B. & Dominh, K. *Geophys. J. R. astr. Soc.* **69**, 15–31 (1982).
18. Cazenave, A., Lago, B. & Dominh, K. *J. geophys. Res.* **88**, 1105–1118 (1983).
19. Cazenave, A. *Earth planet. Sci. Lett.* (in the press).

Siderophile-rich magnetic spheroids from the Cretaceous–Tertiary Boundary in Umbria, Italy

Jan Smit & Frank T. Kyte

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

The magnetic fraction of some Cretaceous–Tertiary (K–T) boundary sediments sampled at two locations in Italy, contains a spheroidal component which has retained some relict mineral textures and, possibly, mineral chemistry from the K–T event. The spheroids contain small crystals (1–50 μm) of magnetite in a highly altered groundmass. These crystals are commonly skeletal, indicating formation by rapid growth from a high-temperature liquid. Variable concentrations of Ni, Mg and Al (differing by at least a factor of 5) in the magnetite may also reflect a relict chemistry. These spheroids contain higher concentrations of Ir and probably other siderophiles than the bulk sediment and may have been an important carrier for these elements at 65 Myr. This discovery supports evidence for a major accretionary event at the end of the Cretaceous.

We have sampled the magnetic fraction of K–T boundary sediments from Furlo and Petriccio in Italy to characterize magnetite-bearing spheroids reported by Montanari *et al.*¹. Three types of microtektite-like spheroidal materials occur in K–T boundary sediments: sanidine-, glauconite-, and magnetite-bearing spheroids¹. The sanidine-bearing spheroids², have oxygen isotopic compositions^{3,4} suggesting authigenic replacement of a precursor material such as glass⁴ or plagioclase¹. Also, Sm–Nd and Rb–Sr isotopic systematics^{4,5} of sanidine spheroids from Caravaca, Spain show that they are distinct from normal detrital sediment and are isotopically consistent with ejecta from an oceanic impact. The magnetite-bearing spheroids are probably related to the sanidine spheroids and provide different clues to the nature of the K–T event.

In these spheroids, the magnetite occurs as small crystals enclosed in a poorly characterized groundmass (Fig. 1a). The magnetite crystals are commonly skeletal and can occur as dendrites (Fig. 1b). These textures, similar to those observed in some igneous rocks^{5,6}, are the result of rapid crystallization from a high temperature liquid. Often the magnetite appears to have crystallized between domains of the enclosing groundmass (Fig. 1a). In this case, entire spheroids may have retained relict



Fig. 1 Polished section images of magnetite-bearing spheroids. *a*, Oil immersion photomicrograph of a spheroid from Furlo, Italy (scale bar, 100 μm). Note that the magnetite (small bright grains) grows between domains of the groundmass. *b*, Backscatter electron image of skeletal magnetite from Furlo which displays well-formed dendrites (scale bar, 100 μm). *c*, Backscatter electron image of skeletal magnetite in a spheroid from DSDP Hole 577A from the Shatsky Rise, North Pacific (scale bar, 10 μm).

Table 1 Microprobe analyses of average spheroid groundmass and typical magnetite grains from Furlo and Petriccio

	Average	Groundmass*		Magnetite			
		1 σ	Range	0-1	0-5	1-2	0-6
SiO ₂	3.84	3.05	1.90-18.71	<0.05	<0.05	<0.05	<0.05
Al ₂ O ₃	5.04	0.80	3.36-7.70	8.42	1.03	4.70	5.16
TiO ₂	3.67	0.47	3.03-4.45	0.60	0.29	0.49	0.53
Cr ₂ O ₃	0.08	0.04	<0.05-0.15	1.51	1.39	0.64	0.57
FeO	47.16	5.23	31.03-54.80	13.90	3.03	7.70	8.32
Fe ₂ O ₃ †				55.03	75.87	70.47	67.75
MnO	1.83	0.28	1.30-2.44	0.28	1.62	0.70	1.16
MgO	0.70	0.29	0.44-1.75	16.33	3.88	10.34	11.50
NiO	0.29	0.17	0.07-0.31	3.89	12.25	5.22	5.42
Total	62.61	4.67	55.74-70.97	99.96	99.36	100.26	100.41

* Groundmass is presented as an average of 38 analyses on 20 spheroids together with typical variation (presented as 1 s.d.), and the entire range of measured concentrations.

† Fe₂O₃ is calculated from measured FeO assuming that Al, Ti and Cr share the trivalent cation site.

textures. Rare faint graphic textures have also been observed in the groundmass.

We suspect that the magnetite is relict compositionally as well as texturally. The magnetite grains have fresh outlines, unlike etched magnetite in highly weathered pillow basalt rims⁷, suggesting that magnetite has been a stable phase in this environment. The magnetites have highly variable compositions (Table 1) although within individual spheroids the compositions are relatively constant. Some crystals have relatively high concentrations of NiO (up to 12%), MgO (up to 17%), or Al₂O₃ (up to 9%) and can best be described as spinel solid solutions with the general formula (Mg, Fe, Ni) (Fe, Al)₂O₄. There is an excellent positive correlation between MgO and Al₂O₃ (Fig. 2), although Mg and Al occupy different cation sites in different proportions. Substitution of Al for the Fe³⁺ cation is as much as 18% and Mg substitution for Fe²⁺ approaches 90%. These values are similar to spinels found in oceanic basalts⁸. However, the K-T magnetites can be distinguished from typical basaltic spinels on the basis of their distinctly higher Ni and lower Cr and Ti contents. Variations in magnetite compositions and textures probably record variations in their chemical environments during crystallization. Important variables may be bulk composition of the molten droplet, cooling rates and crystal nucleation history, and oxygen fugacity.

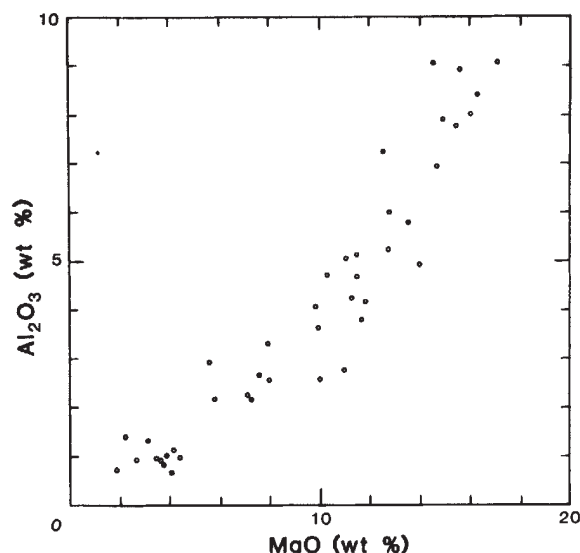


Fig. 2 Plot of Al₂O₃ against MgO measured in magnetites from Furlo and Petriccio, Italy. Note the good positive correlation. Although magnetite grains within single spheroids have relatively constant compositions, there is a wide range of compositions between spheroids.

The spheroid groundmass which encloses the magnetite is highly weathered and probably consists of authigenic minerals. The groundmass is translucent and brownish-coloured and seems to contain hydrous Fe-oxides and clay minerals. This interpretation is supported by microprobe analyses which rarely total >70%, indicating the presence of H₂O and destruction by the electron beam. Iron is by far the most abundant cation in the groundmass, with lesser amounts of Si, Al, Ti and Mn (Table 1). All the variation in groundmass compositions can be attributed to different proportions of the Fe-rich and silicate components or to variable totals. We suspect that the groundmass has re-equilibrated with the surrounding sediments and has not retained a relict major-element chemistry.

The measured Ir concentrations in three magnetic extracts ranged from 24 to 33 ng per g. These are lower than the 69 ng per g reported by Montanari *et al.*¹ (but are considerably higher than the 1-8 ng per g Ir reported for bulk samples of these sites⁹). We attribute this discrepancy to sample heterogeneity. Part of this probably resulted because the magnetic extract contained about 10% of materials other than pure magnetite-bearing spheroids. Differing magnetic extraction techniques may also be partially responsible as the variable magnetite compositions will result in variable magnetic properties. We have observed considerable variation in the Au and Ir concentrations in the three samples analysed (Table 2), and suspect that this is due to sample heterogeneity. One sample was analysed radiochemically for Ir, Pt, Au, Pd, Os and Re. The element/Ir ratios for Pt, Au and Pd were within a factor of 2 of those in CI chondrites. Os was highly depleted; Os/Ir ~0.12 × CI. Re was not detected (<0.8 ng per g) with Re/Ir <0.3 × CI. The roughly chondritic abundances for four of these highly siderophile elements is remarkable for a highly weathered extract that comprises only a small fraction (0.2-0.3 wt%) of the bulk sediment. We estimate that the magnetic fraction presently contains at most 2% of the total siderophiles in the bulk sediment. However, this fraction may originally have been higher if diagenesis caused leaching of siderophiles or destruction of some spheroids.

The only possible sources for these spheroids are an accretionary event or an unusual volcanic event. In a preliminary study of K-T samples from DSDP Hole 577A from the Shatsky Rise, North Pacific, we have found magnetic spherules with skeletal magnetite (Fig. 1c), indicating that this is not a local phenomenon. Because of the wide distribution, high siderophile concentrations and chondritic siderophile ratios, we consider a volcanic source highly unlikely; the source was almost certainly an accretionary event. These spheroids could be derived from a mafic target (for example, oceanic lithosphere) or direct atmospheric formation from a mafic projectile. Although they have some textural similarities to chondritic cosmic spherules¹⁰, their concentration is about 5 × 10⁴ times higher than cosmic spherules in abyssal clays¹¹, and this is an unlikely source for the spheroids. We can also reject ablated dust which could be derived from

Table 2 Siderophile concentrations (ng per g) for magnetic extracts measured by instrumental and radiochemical neutron activation analysis

	Pd	Re	Os	Ir	Pt	Au
Petriccio						
350A	20.3	<0.80	3.10	23.9	29.9	4.58
350B	—	—	—	33.6	—	22
Furlo						
418B	—	—	—	23.6	—	7.5

astronomically induced comet showers^{12–15}, as that would be a fairly long event (probably $>10^5$ yr).

The work was supported by NSF grant OCE 82-08197. The Shatsky Rise sample was supplied by DSDP. We thank W. Lustenhouwer and ZWO/WACOM for some microprobe analyses and A. El Goresy for helpful discussions.

Received 8 February; accepted 1 June 1984.

1. Montanari, A. *et al. Geology* **11**, 668–671 (1983).
2. Smit, J. & Klaver, G. *Nature* **292**, 47–49 (1981).
3. Epstein, S. *Lunar planet. Sci.* **13**, 205–206 (1982).
4. DePaolo, D. J., Kyte, F. T., Marshall, B. D., O'Neil, J. R. & Smit, J. *Earth planet. Sci. Lett.* **64**, 356–373 (1983).
5. Shaw, H. F. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **60**, 155–177 (1982).
6. Haggerty, S. E. *Miner. Soc. Am. Short Course Not.* **3**, 101–300 (1976).
7. Bleil, U. & Smith, B. *Init. Rep. DSDP Leg 51*, 1407–1409 (1979).
8. *Basaltic Volcanism Study Project* (Pergamon, New York, 1981).
9. Alvarez, W., Alvarez, L. W., Asaro, F. & Michel, H. V. *Geol. Soc. Am. Spec. Pap.* **B90**, 305–315 (1982).
10. Blanchard, M. B., Brownlee, D. E., Bunch, T. E., Hodge, P. W. & Kyte, F. T. *Earth planet. Sci. Lett.* **46**, 178–190 (1980).
11. Murrell, M. T., Davis, P. A. Jr, Nishiizumi, K. & Millard, H. T. Jr *Geochim. cosmochim. Acta* **44**, 2067–2074 (1980).
12. Rampino, M. R. & Stothers, R. B. *Nature* **308**, 709–712 (1984).
13. Whitmire, D. P. & Jackson, A. A. IV *Nature* **308**, 713–715 (1984).
14. Davis, M., Hut, P. & Muller, R. A. *Nature* **308**, 715–716 (1984).
15. Alvarez, W. & Muller, R. A. *Nature* **308**, 718–720 (1984).

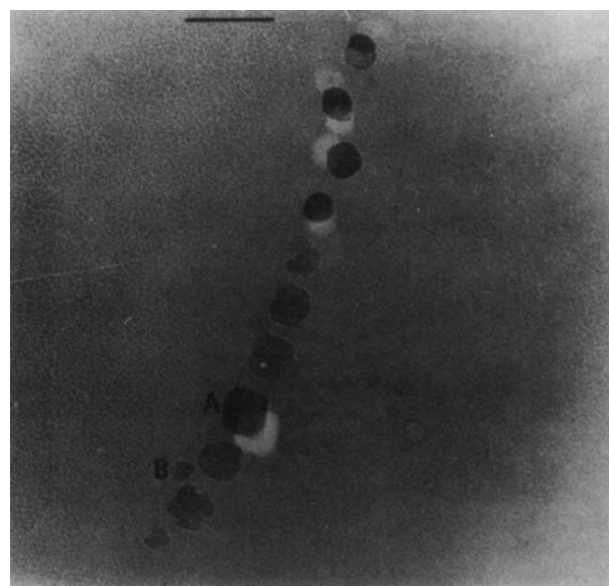


Fig. 1 Chain of magnetite particles imaged within an intact unstained *A. magnetotacticum* cell, showing a combination of large and small particles. Particle A has a characteristic morphology seen in many mature crystals (see Fig. 3). Particle B is irregular in form and appears to be at a different stage in development from particle A. Lattice images of this particle showed crystalline and non-crystalline zones (Fig. 4). The cluster of four particles near to B was also imaged as single crystals with localized amorphous regions. They are spatially separate and therefore not a multi-domain aggregate. Scale bar, 100 nm.

Structure, morphology and crystal growth of bacterial magnetite

Stephen Mann*, Richard B. Frankel†
& Richard P. Blakemore‡

* Inorganic Chemistry Laboratory, University of Oxford,
South Parks Road, Oxford OX1 3QR, UK

† Francis Bitter National Magnet Laboratory, Massachusetts Institute
of Technology, Cambridge, Massachusetts 02139, USA

‡ Department of Microbiology, University of New Hampshire,
Durham, New Hampshire 03824, USA

Recent high-resolution transmission electron microscopy (HRTEM) studies of the structure and morphology of bacterial magnetite (Fe_3O_4) crystals isolated from a magnetotactic coccus¹ and from an unidentified bacterium extracted from sediment² have shown the crystals to be well ordered single-domain particles with a morphology based on a hexagonal prism of {011} faces truncated by specific low index planes. We report here a HRTEM study of intact magnetite crystals (magnetosomes) in the microaerophilic bacterium *Aquaspirillum magnetotacticum*, grown in pure culture^{3,4}. Our aim has been to investigate the structure, morphology and crystal growth of the magnetite particles in the light of a recent Mossbauer spectroscopy study of this organism⁵ which indicated, in addition to magnetite, the presence of hydrated iron(III) oxide phases together with the magnetosomes. Our results show that the mature particles are well ordered single-domain crystals of magnetite with a morphology very different from previously studied crystals and based on an octahedral prism of {111} faces truncated by {100} faces. We also show the first direct evidence for both crystalline and non-crystalline phases within individual magnetosomes. The results are important in aiding elucidation of the crystal growth mechanisms of biogenic magnetite.

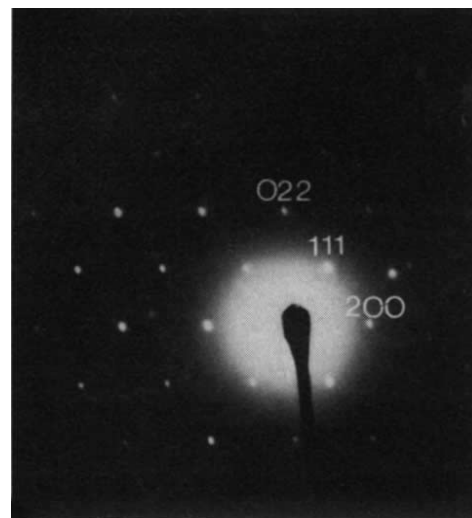


Fig. 2 Selected-area electron diffraction pattern from a mature magnetite crystal of characteristic morphology. The single crystal pattern corresponds to the [011] zone. The (200) and (200) reflections arise from double diffraction. Camera length, 115 cm.

Cells of *A. magnetotacticum* are $\sim 3 \mu\text{m}$ long and contain, on average, 20 intracellular enveloped magnetite particles of diameter 40–50 nm which are organized in a single chain that traverses the cell longitudinally⁶. Cells of *A. magnetotacticum* synthesize magnetite only in microaerobic conditions, accumulating Fe some 20,000–40,000-fold over the extracellular concentration³. The magnetite particles are in the single magnetic domain size range and the chain of magnetosomes imparts a permanent magnetic dipole moment parallel to the cell's axis of mobility such that the bacteria orient in the geomagnetic field⁷.

Intact, unfixed cells were dried down onto carbon-coated electron microscope grids and investigated in a JEOL 200CX high-resolution transmission electron microscope operating at

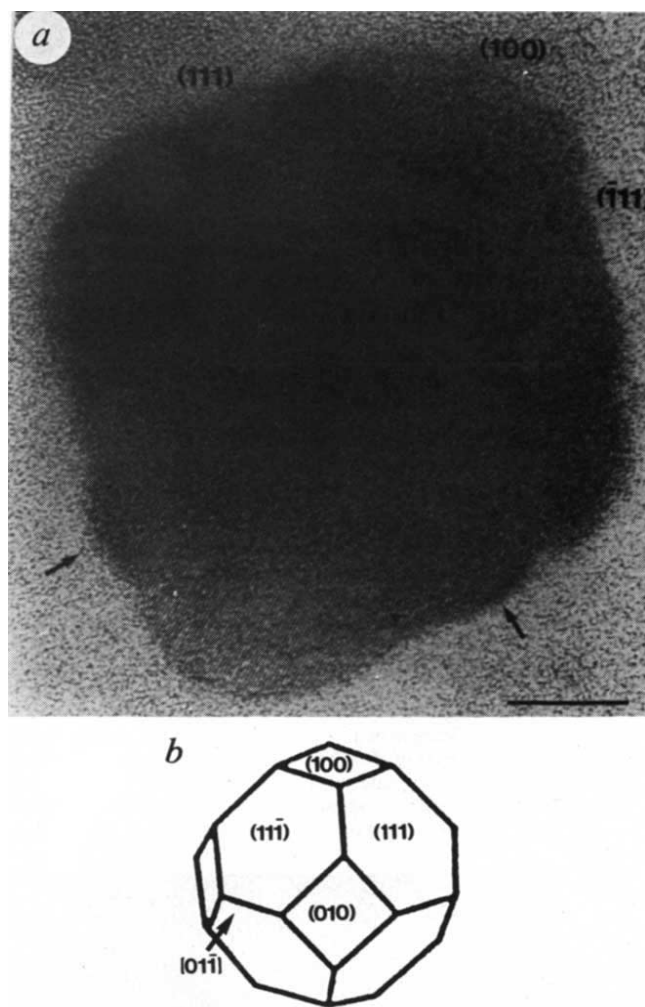


Fig. 3 *a*, HRETEM image of a magnetite crystal imaged in the $[01\bar{1}]$ zone, showing a well ordered single-domain crystal and a characteristic morphology based on an octahedral prism of $\{111\}$ faces truncated by $\{100\}$ faces. The lattice fringes shown correspond to the (022) planes and run perpendicular to the (100) face. Note that the crystal edges are not smooth and show outgrowths on the well developed $\{111\}$ faces (arrow). Scale bar, 10 nm. *b*, Idealized morphology for biogenic magnetite crystals from *A. magnetotacticum*. An octahedral prism of $\{111\}$ faces is truncated by $\{100\}$ faces. Such a crystal form indicates the stabilization of the $\{111\}$ planes over other low index faces such as $\{100\}$.

200 keV with a LaB_6 electron source and a point-to-point resolution of 2.46 Å. Many cells showed chains of preferentially aligned magnetite particles of distinct morphology. Other chains were observed to contain a combination of large, well developed magnetosomes and smaller particles of irregular form, suggesting the presence within the chains of crystals at different stages of development (Fig. 1).

Crystals of distinct morphology (such as crystal A in Fig. 1) showed lattice images consistent with the cubic (Fd3m) inverse spinel structure of magnetite. The lattice images were well defined and run continuously throughout the particles, thus indicating that they are single-crystal domains. Lattice fringes corresponding to the $\{111\}$, $\{222\}$, $\{220\}$, $\{200\}$ and $\{400\}$ planes were imaged and by considering the relative orientation of these fringes, the zone of projection was identified as $[01\bar{1}]$. This was confirmed by selected-area electron diffraction patterns (Fig. 2). The morphology of these mature crystals was inferred from the relative directions of the fringes and crystal edges and is based on an octahedron of $\{111\}$ faces truncated by $\{100\}$ faces (Fig. 3*a, b*). Thus, crystals imaged in the $[01\bar{1}]$ zone often showed edges meeting at characteristic angles of 125° and 110° , corre-

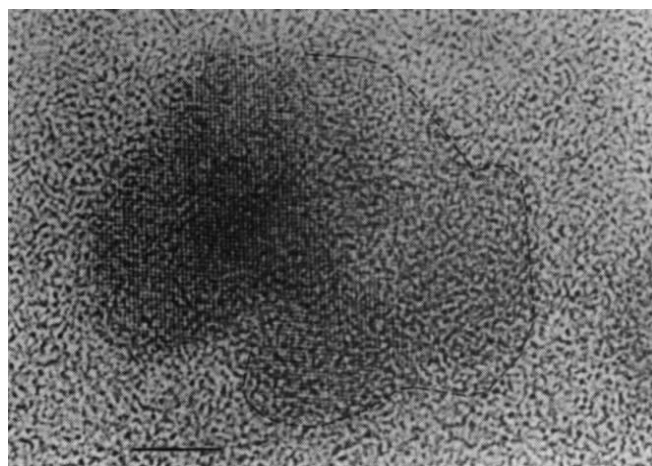


Fig. 4 HRTEM image of particle B of Fig. 1, showing the coexistence of crystalline and non-crystalline phases. The crystalline zone shows well ordered (222) lattice fringes and is a single domain. The fringes extend into the amorphous phase in a preferential direction. The superimposed black dashed line indicates the extent of the low contrast edge of the particle against the background carbon noise of the grid. Scale bar, 5 nm.

sponding to the intersection of (100) and (111) and of $(\bar{1}11)$ and (111) faces. In other mature crystals, these angles were distorted from their theoretical value; this may be a consequence of slight crystal misalignment with respect to the electron beam or it may be a real growth effect.

On the basis of the above morphology, we determined the crystal alignment in bacteria containing chains of well developed crystals. The crystals were found to be preferentially aligned, with the $[111]$ direction parallel to the chain axis. This result is important with regard to magnetotaxis, as the easy axis of magnetization in Fe_3O_4 also lies along the $[111]$ direction.

Our results indicate that the cellular environment in which the crystals grow is often well defined and influences their morphology in a way which is different from that found for other magnetotactic bacteria^{1,2}. Many factors can influence the crystal morphology of biogenic solids⁸, so it may not be surprising that the morphology of the bacterial magnetites seems to be species specific. It is unclear why some mature crystals in *A. magnetotacticum* do not attain the characteristic octahedral morphology; perhaps the crystal growth processes are very susceptible to chemical fluctuations in the intra- and extracellular environments.

Lattice imaging of the irregular particles (for example, particle B in Fig. 1) showed the presence of contiguous crystalline and non-crystalline regions within the magnetosome (Fig. 4). The crystalline zone was always single domain, with well ordered lattice planes of magnetite. No other crystalline phases, such as $\gamma\text{-FeOOH}$, were observed. The lattice fringes often appeared to extend into the amorphous region in a preferential direction. These multi-phase particles probably represent the early stages of magnetite formation, with the non-crystalline material corresponding to the hydrated iron (III) oxide phases identified by Mossbauer spectroscopy³.

On the basis of these results, a mechanism of crystal formation can be offered. The crystallization of bacterial magnetite occurs within a localized region of the cell and involves a non-crystalline precursor of hydrated iron(III) oxide. The lattice images suggest that a growing crystal front of magnetite extends into the amorphous gel. The solid-state rearrangement could occur through a solution front at the interface of the two phases. The final crystal is single domain and often has a characteristic octahedral morphology, indicating that crystal growth is slow and that the $\{111\}$ planes are stable. Because many crystals are preferentially aligned with the $[111]$ direction parallel to the chain axis, both nucleation and growth processes seem to be under biological control. The origin of aqueous Fe^{2+} ions in the crystal growth

mechanism is unclear, although Fe^{2+} ions have been detected in *A. magnetotacticum*⁵. They could arise from partial reduction of the hydrated iron(III) oxide phase at the solid-state interface due to the changes in the local redox potential or they could be transported directly from the cytoplasm or periplasmic space into the crystallizing zone. The adsorption of Fe^{2+} ions at the hydrated iron(III) oxide surface is probably the trigger for magnetite formation¹.

We thank N. Blakemore for help in preparation of the samples. R.B.F. was in part supported by the Office of Naval Research. The Francis Bitter Laboratory is supported by the NSF.

Received 7 March; accepted 1 June 1984.

1. Mann, S., Moench, T. T. & Williams, R. J. P. *Proc. R. Soc. B* **221**, 385–393 (1984).
2. Matsuda, T., Endo, J., Osakabe, N. & Tonomura, A. *Nature* **302**, 411–412 (1983).
3. Blakemore, R. P., Maratea, D. & Wolfe, R. S. *J. Bact.* **140**, 720–729 (1979).
4. Maratea, D. & Blakemore, R. P. *Int. J. Syst. Bact.* **31**, 452–455 (1981).
5. Frankel, R. B., Papaefthymiou, G. C., Blakemore, R. P. & O'Brien, W. *Biochim. biophys. Acta* **763**, 147–159 (1983).
6. Balkwill, D. L., Maratea, D. & Blakemore, R. P. *J. Bact.* **141**, 1399–1408 (1980).
7. Frankel, R. B. & Blakemore, R. P. *J. Magn. magn. Mater.* **15–18**, 1562–1564 (1980).
8. Mann, S. *Struct. Bonding* **54**, 127–174 (1983).

The corticomotoneurone connection is normal in Parkinson's disease

J. P. R. Dick, J. M. A. Cowan, B. L. Day, A. Berardelli, T. Kachi, J. C. Rothwell & C. D. Marsden*

Department of Neurology, Institute of Psychiatry and King's College Hospital Medical School, de Crespigny Park, London SE5 8AF, UK

Voluntary movements in Parkinson's disease are initiated and executed slowly^{1,2}. It is assumed that the motor cortex and its output pathway are intact and that bradykinesia is due to abnormal motor commands delivered to a normal corticospinal system. We have tested this assumption using electrical stimulation of the motor cortex through the scalp in three patients with severe Parkinson's disease, studied during fluctuations from relatively normal mobility when receiving drugs (ON) to severe bradykinesia when not receiving drugs (OFF). Thresholds and latencies for motor cortex stimulation to excite thumb flexor muscles and the resulting fast mechanical responses were the same in both ON and OFF conditions, even though the patients were unable to execute fast thumb flexion movements voluntarily when OFF. We conclude that the excitability and conduction velocity of the corticospinal motor pathways are normal in Parkinson's disease.

Merton and Morton³ demonstrated that brief high-voltage electric shocks to the scalp or spine from a low output resistance stimulator activate visual cortex, motor cortex and spinal cord^{3–5}. Such high-voltage shocks (400 V, giving a peak current of 100–1,000 mA) appear to decrease skin resistance⁶ and cause little local pain when given via a large electrode. Stimulation over the motor cortex produces involuntary muscle contraction at very short latencies, particularly in the distal muscles of the hand and forearm. Activation of thumb muscles appears at the lowest thresholds. A small background voluntary effort in the muscle being tested increases the size of muscle twitch for a given scalp shock⁵.

We have used scalp stimulation of the motor cortex in three volunteer male patients (mean age 48 yr, range 38–59 yr) with Parkinson's disease (mean duration 12 yr, range 8–17 yr; Hoehn and Yahr⁷ severity grade 3). They were selected for their intelligence, severe degree of bradykinesia OFF drugs, and relative lack of tremor; each gave written consent to the study. The patients had been on L-dopa replacement therapy (Madopar, Sinemet and L-dopa) for a mean of 8.6 yr (range 6–13 yr). Each patient exhibited normal mobility with dyskinesias when ON,

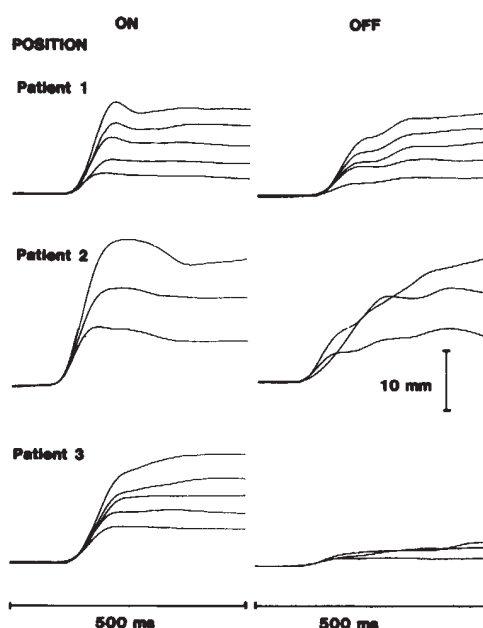


Fig. 1 Thumb flexion movements against a background force of 1.5 N. The amplitude of the movements is shown as mm of displacement of the proximal phalanx of the thumb. The patient was asked to flex his thumb as fast as possible to one of five target points (patients 1 and 3) or to one of three target points (patient 2); these are indicated by the end position of each trace in the ON condition. Averages of 25 trials are shown. Data were collected using a retrospective programme triggered by the upswing of the velocity trace (not shown). The three patients were studied when mobile ON drugs (left), and when bradykinetic OFF drugs (right). There was considerable slowing of movement when OFF drugs. In the OFF condition patient 3 did not achieve the target positions until 4 s had elapsed and for clarity only three traces are shown.

but was wheelchair-bound when OFF. The benefit from each dose of drug tended to last 1–2 h, with peak-dose dyskinesias. None had evidence of cognitive deficit on formal testing. Before the OFF experiment, therapy was withdrawn for at least 24 h (range 24–36 h) and each patient became severely parkinsonian. In two subjects the experiment was done with the patient off medication, then repeated 45–60 min after the administration of L-dopa; in the other subject the experiment was done in the reverse order, first ON then OFF. Patients were deemed ON when general mobility was restored, and rapid flexion movements of their hands did not fatigue over 30 s. The experiment was designed to assess isotonic flexion movements of the thumb whether generated by scalp shock or initiated voluntarily, when in the ON or OFF condition.

The patients were seated with their right, dominant arm immobilized and with their forearm supinated. The elbow and wrist joints were strapped into a plaster of Paris backslab which enclosed the fingers; only the thumb and palm were free. The thumb was securely fitted with a noose around the proximal phalanx which allowed movement at both carpometacarpal and metacarpophalangeal joints. The noose was attached by a light chain to a torque motor (Printed Motors type G9M4H). Thumb position was displayed on a screen visible to the patient as a vertical line, which served as a visual feedback to allow levels of background activation to be kept constant. The electromyogram (EMG) from thenar muscles was pre-amplified by a Devices 3160 amplifier, bandpass-filtered (–3 dB at 80 Hz to 2.5 kHz) and further amplified (Devices 3120). Signals of thumb position (from a potentiometer) and thumb velocity (electronically derived from the position signal) were amplified by a Devices 3160 system, and fed into an on-line PDP 12 computer.

To stimulate the motor cortex, two Ag/AgCl electrodes (surface area 2 cm²) were attached to the scalp, with the cathode at the vertex and the anode 7 cm lateral on a line joining the vertex

* To whom correspondence should be addressed.

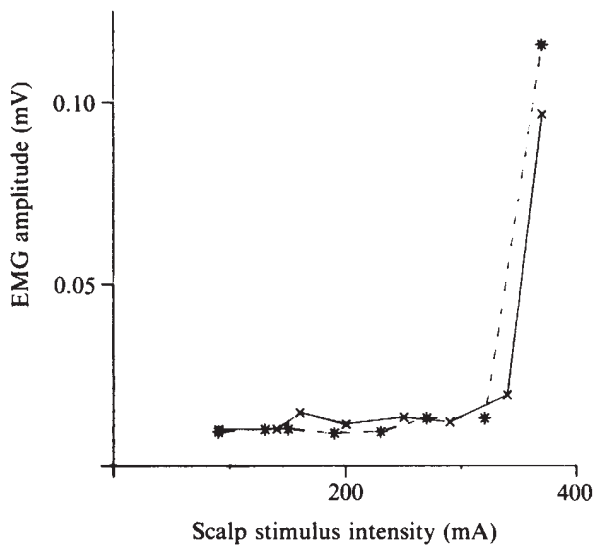


Fig. 2 Threshold scalp stimulus intensities (in mA) required to evoke thenar EMG activation in patient 1 when as relaxed as possible in both ON (solid line) and OFF (broken line) conditions. Each point is the average peak-to-peak size of the EMG potential (in mV) evoked at different stimulus intensities from five scalp shocks. Patients with Parkinson's disease cannot relax completely and there was some background EMG activity measured at lower intensities.

Table 1 Threshold and latency for cortical stimulation in parkinsonian patients either mobile ON or immobile OFF

	Patient 1	Patient 2	Patient 3
Immobile OFF drugs			
Threshold (mA) relaxed	340 (40)	180 (2)	170 (40)
Threshold (mA) contracting	200 (8)	—	130 (8)
Latency (ms) relaxed	22.4 ± 0.3 (17)	20.5 (2)	20.5 ± 0.2 (21)
Mobile ON drugs			
Threshold (mA) relaxed	350 (40)	170 (2)	180 (40)
Threshold (mA) contracting	210 (8)	—	140 (8)
Latency (ms) relaxed	22.9 ± 0.5 (14)	19.6 (2)	20.5 ± 0.2 (36)

Threshold stimulus intensities (in mA) and latency measurements (in ms) for activation of thenar EMG by scalp shocks. Threshold was taken when the EMG trace from the thenar muscles showed a distinct action potential of $>50 \mu\text{V}$ when relaxed (this was often incomplete in patients with Parkinson's disease) and $>0.2 \text{ mV}$ when exerting 1.5 N background force (contracting). The level quoted represents the intensity at which an action potential was observed consistently. In patient 2 the records were inadequate when contracting and no measurement was possible. Latency measurements were taken from rest traces and were measured only from those traces with a distinct onset to the action potential. Temporal resolution was 0.2 ms. Standard errors and the number of records (in parentheses) yielding a measurable quantity are shown.

to the left auditory meatus. Anodal stimulation over the motor hand area with a single short-duration current pulse (time constant of decay $50 \mu\text{s}$) caused involuntary thumb movements. In normal volunteers at rest, the threshold intensity for this reaction was $\sim 400 \text{ mA}$ but, with a small background voluntary effort, thresholds of 140–340 mA were found. The normal latencies for activation of calf, hand and forearm muscles (35 ms, 21 ms and 13 ms, respectively) suggest central conduction velocities of 50 m s^{-1} (ref. 8). It is not known which elements of the motor cortex (Brodmann's area 4) are excited by this method of stimulation. It seems probable that the large efferent neurones of layer V are stimulated with threshold shocks^{9,10}, but whether the axon hillock, initial segment or first internode is activated is unclear^{11,12}.

A set of target positions for voluntary movements was chosen for each patient. A series of rapid voluntary movements to each

target position were recorded and averaged (Fig. 1). The patient was allowed 10–15 practice attempts to gauge the distance of the target and was asked to hold the position once achieved. Encouragement was given for speed and accuracy. We observed a dramatic increase in speed of voluntary movement when the patients were ON, compared with their performance when OFF.

Threshold intensities of scalp shock were determined for each patient at rest and while exerting a background force of 1.5 N. Thresholds were taken when a sharp EMG spike followed by a silence was first seen on the thenar EMG trace. Threshold intensities were similar in both ON and OFF conditions for each patient (Table 1, Fig. 2). The latency of the EMG response also was similar in each patient in both conditions (Table 1). Corticomotoneurone conduction time was calculated in two patients by subtracting an estimate of peripheral nerve conduction time (half the sum of the M-wave latency, which was 3.8

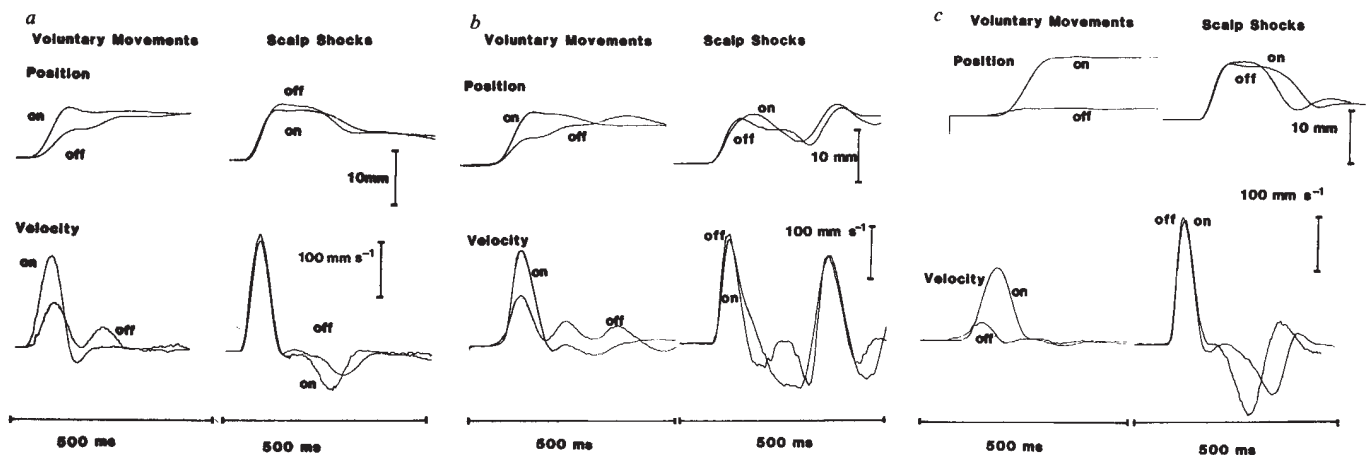


Fig. 3 Averaged position (top) and velocity traces (bottom) from voluntarily initiated thumb movements (left) and movements produced by cortical stimulation (right). The ON and OFF records for the three parkinsonian patients (a, b and c represent patients 1 to 3, respectively) are superimposed in each trace. Scalp shocks were 520, 600 and 600 mA for patients 1–3, respectively; averages of 12, 6 and 14 responses, respectively, are shown for scalp shocks in both conditions. Voluntary movements are taken from Fig. 1 and were chosen to have a target amplitude similar to movements generated by scalp shock. Voluntary movement traces were triggered to start 50 ms before the upstroke of the velocity trace (using a Schmidt trigger). Scalp shock traces were triggered to start 50 ms before the stimulus. The traces demonstrate that when unable to initiate rapid voluntary movements, it is possible to evoke rapid thumb flexion movements by scalp shock. The rapidity of thumb movement generated by scalp shock did not alter when the patient was mobile ON drugs compared with immobile OFF drugs.

and 3.3 ms, respectively, and the F response time, which was 29.8 and 29.3 ms, respectively, following median nerve stimulation at the wrist) from cortical latency. The corticomotoneurone conduction time was comparable to that seen in normal subjects (patient 1, 5.6 ms; patient 3, 4.2 ms) and did not differ between the ON and the OFF condition.

The patients were then asked to exert a 1.5 N force against the torque motor. Suitable movements of the thumb (8–15 mm) were generated by scalp shocks of 520–600 mA. This current intensity was kept constant for each patient and a series (6–14) of movements were recorded. Figure 3a–c shows the averaged position and velocity traces generated by scalp shock in both conditions for each patient. The distance moved, the initial velocity of movement and the duration of the movement were the same whether ON and mobile, or OFF and severely bradykinetic.

Voluntary movements in the ON condition were performed with average initial peak velocities of 174, 176 and 136 mm s⁻¹, somewhat less than those generated by scalp shocks in the same condition (224, 200 and 250 mm s⁻¹). This is because a voluntary movement is achieved by a prolonged burst of EMG activity, whereas a single scalp shock tends to simultaneously recruit all motor units involved¹³. By contrast, voluntary movements were slower in the OFF condition and were accomplished by a series of steps (Figs 1, 3); their initial peak velocities were 87, 93 and 36 mm s⁻¹, considerably slower than those of the movements generated by scalp shocks in the same condition (230, 210 and 255 mm s⁻¹). The movements generated by scalp shocks, however, were similar in terms of initial peak velocity whether the patient was mobile and ON drugs or immobile and OFF drugs. Thus, while all three patients were unable to produce fast movements with a voluntary effort of will when OFF, relatively small cortical shocks could excite spinal cord motor mechanisms, via direct corticomotoneurone connections, to generate fast thumb flexion movements of a size and velocity similar to those seen when ON.

Thus, we have demonstrated that the effects of motor cortex stimulation in patients with moderate to severe parkinsonism are identical whether the patient is dopamine-depleted and OFF or dopamine-replete and ON. Comparable scalp thresholds were found, the latency of muscle activation was constant and the upstroke of the velocity traces was identical. Despite recent discussion of the relevance of dopamine found in the spinal cord¹⁴, these experiments suggest that, when parkinsonian patients are immobile, spinal motor mechanisms can be activated by a corticospinal volley to generate brisk thumb movements with a latency comparable with that of normal volunteers. As the thresholds for scalp stimulation were comparable with normal levels, and as scalp stimulation probably excites the efferent pathway of the motor cortex, it seems likely that the motor cortical output cells are intact in Parkinson's disease. We cannot comment on the intrinsic mechanisms of the motor cortex with this technique, but it seems that the physiological reason for bradykinesia lies upstream of the motor cortical output pathway. The likeliest explanation for bradykinesia is that the movement signal delivered to the motor cortex is defective.

This work was supported by the Research Funds of the Bethlem and Maudsley Hospitals and by the MRC. We thank Mr H. C. Bertoya and Mr R. Miller for technical assistance.

A single gene error of noradrenergic axon growth synchronizes central neurones

Jeffrey L. Noebels

Departments of Neuroscience and Neuropathology,
Children's Hospital Medical Center and
Harvard Medical School, Boston,
Massachusetts 02115, USA

One strategy for deciphering inherited neurological disease is to examine the expression of individual genes controlling the assembly and physiology of specific cell groups within the developing mammalian central nervous system (CNS). This neurogenetic approach, using defined single-locus mutations arising on coisogenic mouse strains, has recently been used to analyse a major class of neuronal membrane diseases involving abnormal excitability, the epilepsies, and to identify examples of hereditary variation in signalling properties at central synapses¹. An interesting mutation, the *Tottering* (*tg*) gene², causes a delayed onset, recessive neurological disorder in the mouse featuring a stereotyped triad of ataxia, intermittent myoclonus and cortical spike-wave discharges accompanied by behavioural absence seizures which resemble petit mal epilepsy³. Axon branches of the locus coeruleus, a noradrenergic brain-stem nucleus, hyperinnervate specific target regions of the *tg* brain⁴. The number of parent coerulean perikarya is unaffected, indicating a true proliferation of the terminal axonal arbor. With the exception of this unusually precise error of axonal growth, no other cytopathology has been identified in the *tg* brain. Here I present evidence that selective lesions of the central noradrenergic axons early in development limit the expression of the disease.

Homozygous *tg/tg* C57BL/6j mice from colonies of the Jackson Laboratory and the Children's Hospital were identified at birth as the progeny of proven homozygous matings. Mutant littermates were given either 6-hydroxydopamine (6-OHDA-HBr, Sigma), 100 mg per kg in 0.05 ml saline with 1 mg ml⁻¹ ascorbic acid, or vehicle alone by subcutaneous injection on the first and/or second postnatal days. Electroencephalographic recordings of freely moving mice 18–28 days of age were obtained daily during random 1–2-h periods. Mice were killed at 3–5 months old and their brains excised for regional examination of glyoxylic acid-induced histofluorescence⁴ and catecholamine content assayed by HPLC–EC with dihydroxybenzylamine as an internal standard to a sensitivity of 20 pg (ref. 5).

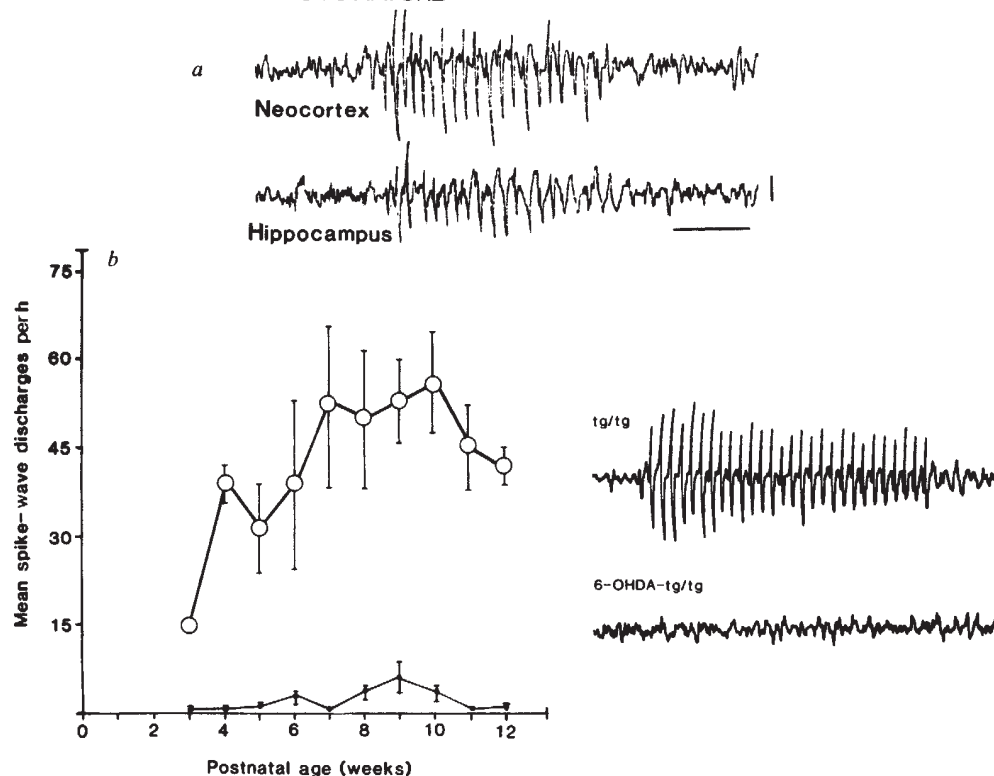
The reproducible neuroanatomical lesion resulting from systemic neonatal injection of 6-OHDA has been carefully examined in rodent brain^{6–8} and describes accurately the changes observed in C57BL/6j *tg/tg* mice injected identically in the present study. The salient finding is a selective, permanent and near-total degeneration of distal catecholamine axons within forebrain target areas innervated by the locus coeruleus (LC), notably neocortex and hippocampus. LC axons within brain-stem and cerebellar regions more proximal to the parent nucleus escape permanent injury. Low doses of neurotoxin generally spare other CNS neurotransmitter systems⁹ and cortical cytoarchitecture¹⁰.

The earliest difference detected in treated *tg/tg* mutants was the absence of generalized, non-convulsive cortical spike-wave discharges. These spontaneous bursts (300–400 μ V amplitude, 6–7 spikes per s, 1–10 s duration) are bilaterally symmetrical and appear in the hippocampus in synchrony with the overlying neocortex (Fig. 1a). Synchronous discharges can be detected infrequently (1–2 h⁻¹) in awake *tg/tg* mice by postnatal day 18, increasing nearly 30-fold in the ensuing 2 weeks to mean adult rates of 28–60 spike bursts per h (mean 47.0, $n=24$; Fig. 1b). Neonatal subcutaneous injections of 6-OHDA entirely prevented the appearance of cortical discharges and behavioural

Received 13 April; accepted 29 May 1984.

1. Parkinson, J. *An Essay on the Shaking Palsy* (Sherwood, Nealy and Jones, London, 1817).
2. Draper, I. T. & Johns, R. J. *Bull. Johns Hopkins Hosp.* 115, 465–480 (1964).
3. Merton, P. A. & Morton, H. B. *Nature* 285, 227–228 (1980).
4. Merton, P. A. & Morton, H. B. *J. Physiol., Lond.* 305, 9–10P (1980).
5. Merton, P. A., Hill, D. K., Morton, H. B. & Marsden, C. D. *Lancet* ii, 597–600 (1982).
6. Hill, D. K., McDonnell, M. J. & Merton, P. A. *J. Physiol., Lond.* 300, 2–3P (1980).
7. Hoehn, M. M. & Yahr, D. *Neurology, Minneap.* 17, 427–442 (1967).
8. Marsden, C. D., Merton, P. A. & Morton, H. B. *J. Physiol., Lond.* 328, 6P (1982).
9. Phillips, C. G. *Q. J. exp. Physiol.* 41, 70–84 (1956).
10. Asanuma, H. & Sakata, H. *J. Neurophysiol.* 30, 35–54 (1967).
11. Phillips, C. G. & Porter, R. *Corticospinal Neurons*, 84–85 (Academic, New York, 1977).
12. Landau, W. M., Bishop, G. H. & Clare, M. H. *J. Neurophysiol.* 28, 1206–1222 (1965).
13. Marsden, C. D., Merton, P. A. & Morton, H. B. *J. Physiol., Lond.* 312, 5–6P (1980).
14. Lindvall, O., Bjorklund, A. & Skagberg, G. *Ann. Neurol.* 14, 255 (1983).

Fig. 1 *a*, Simultaneous uni-lateral surface and depth recordings of spontaneous 6–7 s⁻¹ spike and wave burst in adult *tg/tg* mutant neocortex (above) and hippocampus using a symmetrical array of Teflon-coated Ag wire electrodes (0.001 inch bare diameter) implanted bilaterally overlying the dura, and within the hippocampal formation. Each discharge is always accompanied by a simultaneous episode of behavioural arrest, with a staring posture and occasional myoclonic jerk of limb or jaw. A time series evaluation of discharge intervals in untreated *tg* mice monitored continuously over 24 h revealed that seizures are predictably absent only during sleep, and depart significantly in distribution from a Poisson-like process. Calibration bars, 1 s, 10 μ V. *b*, Left, effect of neonatal systemic exposure to 6-OHDA on spontaneous expression of cortical spike-wave discharges in developing *tg/tg* mice (closed circles, *n* = 9), compared with saline-injected control *tg/tg* littermates (open circles, *n* = 9). Early and permanent suppression of the inherited seizures was absolute in five out of nine 6-OHDA-treated mutants. In the final week, the control group averaged 41.4 discharges per h (± 3.6 s.e.m.) compared with 0.1 discharges h⁻¹ in the treated group. Right, upper, synchronous cortical spike burst in untreated *tg/tg* adult mouse. A typical mutant averages 900–1,000 bursts per day. Lower, representative tracing of the unaltered background EEG activity in littermate *tg/tg* treated with systemic 6-OHDA at birth. This mutant failed to show a single burst discharge in over 100 h of daily monitoring.



absence seizures, without otherwise altering background electroencephalogram (EEG) rhythms of freely behaving mice (Fig. 1b).

In five of nine treated *tg/tg* mice, each from different litters, the effect of the lesion was absolute; no spike-wave discharges were ever observed throughout the longest period (12 weeks) of daily monitoring. Histofluorescence examination of LC axon terminal innervation fields revealed a virtually complete absence of fluorescent axons in all laminae throughout neocortex and hippocampus (Fig. 2). Less distinct noradrenergic alterations in thalamic, brain-stem and cerebellar nuclei were noted, and the fluorescence intensity of LC somata did not diminish. High-pressure liquid chromatography-electrochemical detection (HPLC-EC) determinations (*n* = 5) of regional noradrenaline levels revealed decreases in the neocortex (10.8 ± 2.6 (s.e.m.) ng per g) and hippocampus (12.5 ± 5 ng per g) to 3–6% of tissue levels in untreated *tg/tg* littermates (361.2 ± 31 ; 403.2 ± 83 ng per g, respectively). Whole cerebellar noradrenaline content (296 ± 43 ng per g) remained at 98.6% of untreated adult levels (302 ± 59 ng per g). In the remaining four mutants, all from one litter, discharges were initially absent, but appeared later (5–8 weeks) on isolated occasions at a greatly decreased rate (mean 1–2 h⁻¹), duration (1–5 spikes) and amplitude (<200 μ V), similar to those observed in the youngest (P18) untreated *tg/tg* mice. Histofluorescence surveys in these mice showed scattered noradrenergic axon patches in all neocortical layers, suggesting that the original lesion was incomplete, although the elapsed time is compatible with partial regrowth of the LC axonal arbor^{6,8}.

The second result of the neonatal neurotoxin injection was a reversal of the hereditary gait disturbance. Ataxia is the earliest overt trait of the homozygous *tg* mouse, evident by 4 weeks of age². The pathognomonic gait, an outward splaying of the hindlimbs at rest and broad-based, halting steps, was visibly diminished by early noradrenergic axon lesions. Adult *tg/tg* mice given neonatal injections never tottered, but ran rapidly

with even, unhesitating steps, only barely distinguishable from normal. Measurements of inked footprints revealed a consistently shorter stride on a narrower base. The protective effect of neonatal 6-OHDA lesions on ataxia did not diminish over the 5-month experimental period.

The third component of the triad—paroxysmal focal limb myoclonus—was variably present in neonatally treated *tg/tg* mice, although the few episodes observed were reduced in duration and severity. The effect was not explored quantitatively owing to the rare spontaneous occurrence of the event. Note that neocortical synchronization is not a characteristic feature of the myoclonic episode and that diencephalic and brain-stem nuclear regions so far recognized to be actively involved in the disturbance (zona incerta, nuclei parafascicularis, ruber and pontis oralis³) were not denervated by the systemic lesion.

These findings link aberrant regional LC hyperinnervation to the final neurological expression of the *tg* gene, but do not resolve whether the failure to express *tg* traits is a simple consequence of lowered catecholamine release onto target cells or an indirect result of developmental reorganization following early loss of LC afferents. Neonatal noradrenergic denervation may disturb normal maturation of CNS synapses by destabilizing target membrane receptor populations¹¹, altering adenylate cyclase-linked metabolism¹² or by triggering growth of neighbouring non-adrenergic terminals¹³.

To determine whether reduction of the LC innervation could acutely reverse abnormal cortical synchronization once the morphological and behavioural traits of the *tg* gene were fully expressed, adult *tg/tg* mice were injected unilaterally in the LC nucleus. The extent of LC ablation was subsequently evaluated by fluorescence microscopy of pontine sections. In three mice with verified coeruleus lesions, both the rate and duration of discharges increased up to sevenfold over the initial 48 h, then decreased, coincident with the period of axonal release and cortical noradrenaline depletion¹⁴. Over a 10-week period, mice showed a mean 82% decrease in EEG discharge rate (12 h⁻¹)

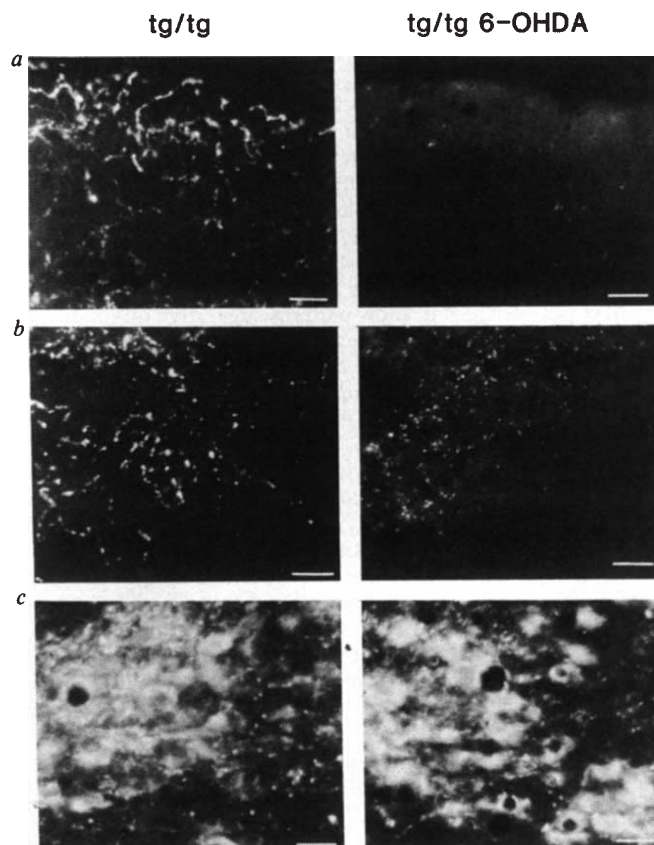


Fig. 2 Regional fluorescence histochemistry of LC noradrenaline innervation patterns within the adult *Tottering* central nervous system (*tg/tg*) compared with littermate treated on postnatal day 1 with a subcutaneous injection of 6-OHDA. The abnormally profuse distal LC axon projection to the neocortex (*a*) and hippocampal dentate gyrus (*b*) degenerates to less than 10% of the total noradrenaline content following neonatal neurotoxin exposure. Gold autofluorescent granules in the hippocampus persist in the absence of green noradrenergic fibre fluorescence. Parent neurones of the locus coeruleus (*c*), along with the proximal axons in brain stem and cerebellum, retain normal fluorescent activity. Scale bars, 25 μ m.

in the final session compared with the pretreatment rate (66 h^{-1}); 75% of the suppression occurred in the first week. Isolated spikes and asymmetries in the onset and termination of discharges occurred; in two mutants, discharges were often restricted to the contralateral hemisphere (Fig. 3). These patterns are never observed in untreated mice. Histochemistry comparisons showed asymmetrical but not total noradrenergic fibre depletion within the ipsilateral neocortex and hippocampus.

These experiments relate the inherited expression of two discrete neurological disturbances, spike-wave seizures and ataxia, to a gene-linked abnormality within a single neuromodulatory cell system of the pontine brain stem which projects axons to widespread target regions of the CNS. Note that the postnatal growth of LC axon terminal collaterals peaks 3–5 weeks after birth¹⁵, coinciding with the first signs of ataxia and seizures in the *Ig* mouse. Although the specific neuronal pathways underlying each of these highly integrated disorders are unknown, the data show that the abnormal excitability is directly influenced by the balance of noradrenergic innervation, and further, that there is no critical organizational stage in postnatal CNS development which irreversibly commits the affected neurones to altered firing patterns.

The link between LC hyperinnervation and the development of spontaneous cortical discharges underlying epilepsy in the *tg* mouse provides new evidence that the locus coeruleus can

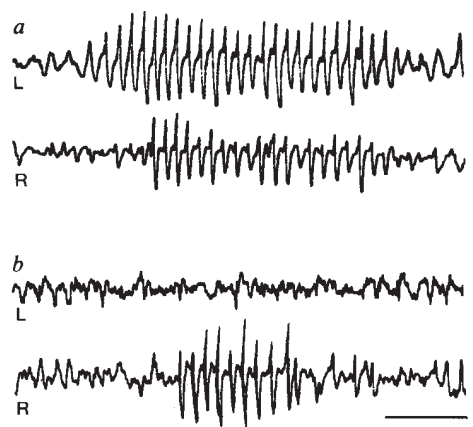


Fig. 3 *a*, Asymmetry in onset after 3 days; *b*, unilateral suppression of synchronous discharges in left cortical hemisphere of adult *tg/tg* 1 week after 6-OHDA injection into the region of the left locus coeruleus ($4 \mu\text{g}$ in $0.25 \mu\text{l}$ 0.9% saline with $0.2 \mu\text{g ml}^{-1}$ ascorbic acid; stereotaxic surface coordinates AP -3.0 mm , 1 1.3, V -3.0 via 32-gauge needle under avertin anaesthesia). Bilateral discharges were also seen in this mouse at a greatly reduced rate. Lateralized discharges are never seen in untreated mutants. Calibration bars, 1 s, 100 μV .

facilitate synchronous bursting in central neurones. Several actions of noradrenaline on membrane currents suggest a basis for this modulatory behaviour. For example, noradrenaline hyperpolarizes and reduces spontaneous impulse activity in principal outflow neurones of neocortex, hippocampus, thalamus and cerebellum while enhancing excitatory and inhibitory postsynaptic responses to synaptic input^{16–18}. In hippocampal neurones, this increase in the synaptic transfer function is partly mediated by a blockade of membrane accommodation; noradrenaline-induced decrements in a Ca^{2+} -sensitive K^{+} conductance prolong action potential repolarization, resulting in bursting rather than solitary impulse firing during pulse depolarization^{19,20}. Recruitment of neurones into the hyper-synchronous discharge by noradrenaline may depend upon the individual cellular profile of membrane receptor subtypes²¹, and the local pattern of LC innervation²². It seems significant that the desynchronizing effect of noradrenaline depletion in inherited spike-wave absence seizures differs from the proconvulsant action observed in experimental models of generalized motor seizures^{23,24}, underlining a further pharmacological and perhaps fundamentally physiological distinction between these two major forms of epilepsy.

I thank Tom Fox, Ellen Grass and Richard Sidman for their assistance. This work was supported by the Klingenstein Foundation.

Received 18 October 1983; accepted 4 May 1984.

- Noebels, J. L. *Fedn Proc.* **38**, 2405–2410 (1979).
- Green, M. C. & Sidman, R. L. *J. Hered.* **53**, 233–237 (1962).
- Noebels, J. L. & Sidman, R. L. *Science* **204**, 1334–1336 (1979).
- Levitt, P. & Noebels, J. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4630–4634 (1981).
- Mefford, I. N., Gilberg, M. & Barchas, J. D. *Analyt. Biochem.* **104**, 469–472 (1980).
- Sachs, Ch., Pycock, C. & Jonsson, G. *Med. Biol.* **52**, 55–65 (1974).
- Levitt, P. & Moore, R. Y. *Anat. Embryol.* **158**, 133–150 (1980).
- Schmidt, R. H., Kasik, S. A. & Bhatnagar, R. K. *Brain Res.* **191**, 173–190 (1980).
- Jacks, B. R., DeChamplain, J. & Cordeau, J.-P. *Eur. J. Pharmacol.* **18**, 353–360 (1979).
- Lidov, H. G. W. & Molliver, M. E. *Dev. Brain Res.* **3**, 81–108 (1982).
- Wolfe, B. B., Minneman, K. P. & Molinoff, P. B. *Brain Res.* **234**, 474–479 (1982).
- Walton, K. G., Miller, E. & Baldessarini, R. J. *Brain Res.* **177**, 512–522 (1979).
- Amaral, D. G., Avendano, C. & Cowan, W. M. J. *Comp. Neurol.* **194**, 171–191 (1980).
- Sachs, C. & Jonsson, G. in *Chemical Tools in Catecholamine Research* Vol. 1 (eds Jonsson, G., Malmfors, T. & Sachs, Ch.) 163–171 (North-Holland, Amsterdam, 1975).
- Elias, M., Deacon, T. & Caviness, V. S. Jr. *Brain Res.* **233**, 652–656 (1982).
- Rogawski, M. A. & Aghajanian, G. K. *Nature* **287**, 731–734 (1980).
- Waterhouse, B. D., Moises, H. C. & Woodward, D. J. *Expl. Neurol.* **69**, 30–49 (1980).
- Moises, H. C. & Woodward, D. J. *Brain Res.* **182**, 327–344 (1980).
- Madison, D. V. & Nicoll, R. A. *Nature* **299**, 636–638 (1982).
- Jefferys, J. G. R. & Haas, H. L. *Nature* **300**, 448–450 (1982).
- Szabadi, E. *Neuropharmacology* **18**, 831–843 (1979).
- Morrison, J. H., Molliver, M. E., Grzanna, R. & Coyle, J. T. *Neuroscience* **6**, 139–158 (1981).
- Maynert, E. W., Marczyński, T. J. & Browning, R. A. *Adv. Neurol.* **13**, 79–147 (1975).
- Corcoran, M. E. & Mason, S. T. *Brain Res.* **190**, 473–484 (1975).

Molecular evidence for new mutation at the *hprt* locus in Lesch-Nyhan patients

Thomas P. Yang*§, Pragna I. Patel*§, A. Craig Chinault*†§, John T. Stout†, Laird G. Jackson||, Beth M. Hildebrand§ & C. Thomas Caskey*†‡§¶

Departments of Medicine*, Biochemistry† and Cell Biology‡, Baylor College of Medicine and Howard Hughes Medical Institute§, Houston, Texas 77030, USA and Division of Medical Genetics||, Jefferson University, Philadelphia, Pennsylvania 19107, USA

Hypoxanthine-guanine phosphoribosyltransferase (HPRT; EC2.4.2.8), which functions in the metabolic salvage of purines, is encoded by an X-linked gene in man. Partial HPRT deficiencies are associated with gouty arthritis, while absence of activity results in Lesch-Nyhan syndrome (L-N). L-N patients fail to reproduce and the heterozygous state appears to confer no selective advantage¹. Thus, Haldane's principle² predicts that new mutations at the *hprt* locus must occur frequently in order for L-N syndrome to be maintained in the population. This constant introduction of new mutations would be expected to result in a heterogeneous collection of genetic lesions, some of which may be novel³. As we report here, the mutations in the *hprt* gene of seven L-N patients, selected from an initial survey of 28 patients, have been characterized and all were found to be distinctly different, as predicted. The origin of one unusual mutation has been identified by analysis of DNA from four generations of family members. Further molecular analysis of the origin of new mutations at the *hprt* locus should aid in resolving the issue of an apparent difference in the frequency of *hprt* mutations in males and females³⁻⁵.

¶ To whom correspondence should be addressed.

DNA from 28 unrelated L-N patients was digested with four different restriction enzymes (*Bam*HI, *Bgl*II, *Pst*I, *Msp*I) and analysed by Southern blotting using a labelled full-length *hprt* cDNA^{6,7} probe. Twenty-three DNA samples possess Southern blot patterns identical to that found among normal controls. These include five patients who display at least one of two X-linked *Bam*HI restriction fragment length polymorphisms⁸. Thus, 83% of L-N patients surveyed did not show major *hprt* gene alterations and are presumed to have mutations that are undetectable by these Southern blot analyses.

Five (18%) of the patients screened, including one with a *Bam*HI polymorphism, show patterns suggestive of major gene alterations. The results are summarized in Table 1 and illustrated in Fig. 1. Analysis of these alterations is made possible by comparison with both the mouse and normal human *hprt* gene structures. The mouse gene is composed of nine exons distributed over 33 kilobases (kb) of DNA⁹. The human gene is at least as large, with the first three and the last three exons showing intron-exon junctions identical to those of the corresponding exons in the mouse gene (P.I.P. *et al.*, in preparation). Thus, our preliminary characterization of human L-N mutations is based on the assumption that the structure of the remaining human exons will be identical to those in the mouse gene. Figure 1a shows the exon assignments for the genomic Southern blot patterns of a control *Pst*I digest. Similar exon correlations have been performed for DNA fragments generated by digestion with *Bam*HI, *Bgl*II, *Hinc*II, *Taq*I, *Eco*RI and *Hind*III.

By examining the alteration or absence of normal genomic restriction fragments which correspond to specific exons in L-N Southern blots, it is possible to predict the nature and regions of the L-N gene alterations (Fig. 2). All three of the patients with gene deletions (RJK 853, RJK 849, GM 3467) show cytogenetically normal X chromosomes (D. Ledbetter, unpub-

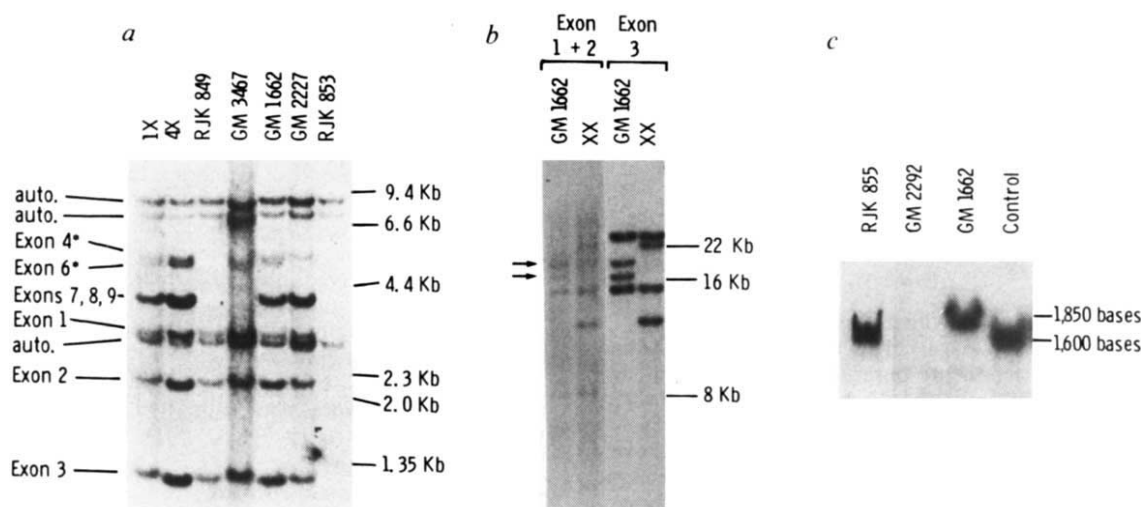


Fig. 1 Representative Southern and Northern blot patterns of normal and Lesch-Nyhan DNA. *a*, *Pst*I-digested genomic DNA from normal male (1X), female with 4 X chromosomes (4X) and L-N patients (RJK 849, GM 3467, GM 1662, GM 2227, RJK 853) was hybridized with a labelled full-length *hprt* probe. Restriction fragments from the X-linked, expressing *hprt* gene are shown with corresponding exon elements contained within these fragments. Asterisks indicate ambiguous assignment of exon 5 with exon 4 or exon 6 bands. "Auto." indicates autosomally-linked fragments¹⁰. *b*, *Bam*HI-digested DNA from a normal female (XX) and GM 1662, hybridized to an *hprt* cDNA subfragment probe containing only exons 1 and 2 and a probe containing only exon 3. Both probes hybridize to two aberrant fragments (17 kb, 19 kb) indicated by the arrows. Southern blot analysis of exon 1 suggests no alteration of the 5' end of the gene (data not shown). In addition, genomic *Pst*I bands in *a* representing exons 2 and 3 in GM 1662 are relatively more intense than other X-linked bands (a twofold increase in intensity was confirmed by densitometry; data not shown). The simplest interpretation of these data, assuming a single mutational event, is a partial gene duplication involving exons 2 and 3. For Southern blots, 5 µg of genomic DNA (from cultured lymphoblast or fibroblast cell lines) per lane was digested to completion with restriction enzymes, fractionated on 0.8% agarose gels and transferred to nitrocellulose filters. Hybridization with labelled probes was carried out by slight modification of published procedures²⁰. Cloned full-length *hprt* cDNA or subfragments were labelled with [α -³²P]dCTP (Amersham) by nick-translation (Bethesda Research Laboratories kit). cDNA subfragments were generated by restriction enzyme digestion and recovered from low-melting agarose²¹. Washing of hybridized filters was performed at 65 °C in 2×SSC, 0.1% SDS and autoradiography was done at -80 °C with Dupont Cronex intensifying screens. A detailed description of the correlation between *hprt* cDNA exons and genomic Southern blot bands of normal DNA shown above will be presented elsewhere (P.I.P. *et al.*, in preparation). *c*, Representative Northern blots of *hprt* mRNA from L-N patients. Hybridization was performed with nick-translated full-length *hprt* cDNA. RNA preparation, electrophoresis and hybridization were carried out as described by Fuscoe *et al.*²². Northern analysis was performed on 50 µg of total cellular RNA.

lished results). GM 1662 represents a potentially novel alteration at the *hprt* locus and is described in Fig. 1b.

Figure 3 summarizes a family study carried out to detect the origin of the observed gene alteration in GM 1662. DNA was obtained from the maternal grandmother, mother, sister and niece of the patient and analysed by digestion with *Bgl*II, Southern blotting and hybridization with labelled full-length cDNA. An aberrant band at 4.1 kb, characteristic of the mutation in GM 1662, was used as the marker for carriers of this mutation. As shown in Fig. 3a, only the maternal grandmother lacks this aberrant band. Therefore, this mutation is likely to have originated in the germ line of the maternal grandmother or grandfather. We believe this to be the first example of detection and characterization of a new mutation in man at the molecular level.

Characterization of messenger RNA from L-N patients has also been performed by Northern blot analysis (Fig. 1c). None of the deletion patients nor GM 2227 produces a detectable, stable mRNA (data not shown). GM 1662 produces an *hprt* mRNA slightly larger than normal which is consistent with the proposed partial gene duplication. Also shown are Northern blots of two additional L-N patients, RJK 855 and GM 2292, both of whom display normal *hprt* Southern blot patterns. RJK 855 produces a normal-sized mRNA at normal levels, and GM 2292 produces no detectable *hprt* mRNA. The mutation in RJK 855 is therefore distinct from the lesion in GM 2292.

Our studies describe seven independent mutations at the *hprt* locus resulting in L-N syndrome, each with a different molecular alteration. Three of these L-N *hprt* gene alterations seem to be similar in nature to deletions previously described in thalassaemia patients¹⁰⁻¹⁴ and in X-linked factor IX-deficient haemophilia patients¹⁵. Many of the L-N *hprt* mutations which

do not show major gene alterations are likely to be similar to those found in thalassaemias, such as base substitutions¹⁶, which in turn, may lead to aberrant splicing¹⁷ or to the introduction of nonsense codons¹⁸. However, the type of mutation seen in GM 1662, and perhaps GM 2227, may be unique, thus far, to *hprt*.

That each of the seven defective genes analysed is distinctly different from one another is consistent with the prediction that the maintenance of X-linked recessively inherited diseases by new mutations will be reflected by molecular heterogeneity¹. This is also suggested by earlier studies of *hprt* protein and enzymatic parameters¹. Furthermore, our studies emphasize the potential use of cloned probes to detect and study new mutations in man. Francke *et al.*³ find the incidence of new mutations in L-N patients to be lower than that predicted by Haldane² and therefore suggest that the frequency of mutation at the *hprt* locus may be different in males and females. Resolution of this question would be aided by the use of cloned polymorphic probes closely linked to *hprt* for developing molecular haplotypes. This would permit determination of the parental origin of new mutations in cases where the mutation is first detectable in a carrier mother such as the one described here. In addition, preferential association of certain types of germ-line mutations with males or females could be ascertained.

We thank R. DeMars, R. McInnes, P. Benke, R. Howell, W. Nyhan, H. Ropers, R. Cox and W. Crowder for providing cultured L-N cells or blood. T.P.Y. was supported by NIH training grant GM07526 and NIH postdoctoral fellowship F32 GM09753, P.I.P. by the Robert A. Welch Foundation and an Arthritis Foundation Fellowship, A.C.C. and C.T.C. by the Howard Hughes Medical Institute and NIH grant AM31428.

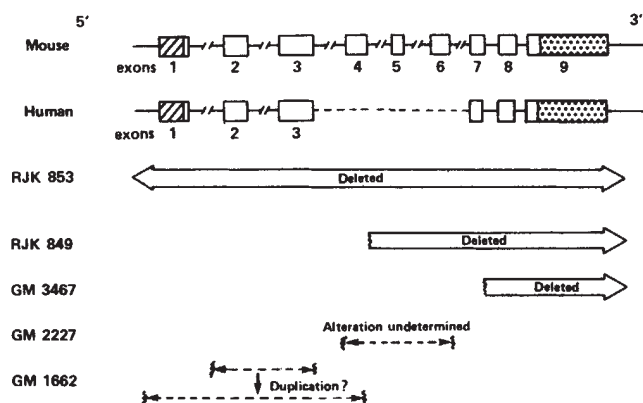


Fig. 2 Summary of altered *hprt* gene in Lesch-Nyhan patients. Arrows indicate the region of the gene alteration. The endpoints of the alterations are approximate. Hatched areas show 5'-untranslated region; stippled areas show 3'-untranslated region. In patient RJK 853, all X-linked genomic Southern blot bands were absent and no new bands were visible. Thus, this patient is presumed to possess a total deletion of the *hprt* gene. In L-N patient RJK 849, all genomic Southern blot bands containing exons 7, 8 and 9 and nearly all bands associated with exons 4, 5 and/or 6 were absent. However, bands containing exons 1, 2 and 3 were always present, with no new aberrant bands detectable. Therefore, we conclude that this patient has a partial gene deletion at the 3' end of the gene which includes exons 7, 8 and 9 and extends into the region coding for exons 4, 5 and/or 6. Similar analysis of GM 3467 suggests a partial gene deletion beginning within the region coding for exons 7, 8 and 9 and extending 3' to the last exon. This latter prediction was confirmed by hybridization of a cDNA subfragment containing only a portion of exon 9 to GM 3467 DNA. This probe did not hybridize to *Bgl*II and *Pst*I bands known to contain exon 9 and no new bands were seen (data not shown). GM 2227 displays Southern blot alterations which are associated with bands representing exons 4, 5 and/or 6 while no alterations are found in exons 1, 2, 3, 7, 8 and 9. However, the exact nature of this alteration has not been determined. See Fig. 1b legend for a description of GM 1662.

Table 1 Summary of Southern blot analysis of Lesch-Nyhan patients showing gene alterations

Patient	Enzyme detecting an alteration			
	<i>Bam</i> HI	<i>Bgl</i> II	<i>Pst</i> I	<i>Msp</i> I
RJK 849	+		+	
GM 3467	+	+	+	
GM 2227		+	+	+
GM 1662	+	+		
RJK 853	+	+	+	+

+ Indicates a restriction enzyme which detects an alteration. GM 1662, GM 3467, GM 2227 are from the Human Genetic Mutant Cell Repository, Camden, New Jersey, RJK 849 from R. DeMars and RJK 853 from R. McInnes.

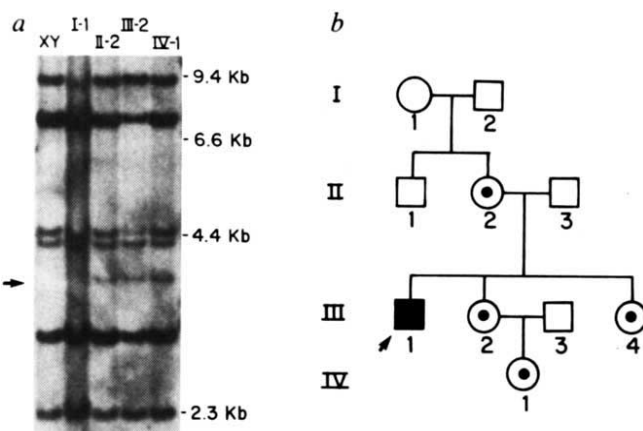


Fig. 3 Molecular analysis of *hprt* gene alteration in the family of patient GM 1662. *a*, Southern blot of *Bgl*II-digested DNA from normal male (XY), maternal grandmother (I-1), mother (II-2), sister (III-2) and niece (IV-1) of proband. The arrow indicates a 4.1-kb *Bgl*II band associated with the gene alteration in GM 1662. *b*, Pedigree of GM 1662 family showing occurrence of new mutation. $\circ$, Non-carrier female; $\odot$, carrier female; $\square$, unaffected male; $\blacksquare$, affected male. The proband is indicated by an arrow.

Received 21 February; accepted 16 May 1984.

- Wilson, J. M., Young, A. B. & Kelley, W. N. *New Engl. J. Med.* **309**, 900-910 (1983).
- Haldane, J. B. S. *J. Genet.* **31**, 317-326 (1935).
- Franke, U. *et al. Am. J. hum. Genet.* **28**, 123-137 (1976).
- Franke, U. *Adv. exp. Med. Biol.* **154**, 18-21 (1982).
- Morton, N. E. & Lalouel, J. M. *Am. J. hum. Genet.* **29**, 304-307 (1977).
- Brennand, J., Chinault, A. C., Konecki, D. S., Melton, D. W. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1950-1954 (1982).
- Jolly, D. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 477-481 (1982).
- Nussbaum, R. L., Crowder, W. E., Nyhan, W. L. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4035-4039 (1983).
- Melton, D. W., Konecki, D. S., Brennand, J. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2147-2151 (1984).
- Ottolenghi, S. *et al. Nature* **251**, 389-391 (1974).
- Taylor, J. M. *et al. Nature* **251**, 392-393 (1974).
- Frisch, E. F., Lawn, R. M. & Maniatis, T. *Nature* **279**, 598-603 (1979).
- Flavel R. A., Bernards, R., Kooter, J. B. & De Boer, E. *Nucleic Acids Res.* **6**, 2749-2760 (1979).
- Orkin, S. H., Kolodner, R., Michelson, A. & Hussan, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3558-3562 (1980).
- Gianelli, F. *et al. Nature* **303**, 181-182 (1983).
- Orkin, S. H. *et al. Nature* **296**, 627-631 (1982).
- Treisman, R., Proudfoot, N. J., Shander, M. & Maniatis, T. *Cell* **29**, 903-911 (1982).
- Chang, J. C. & Kan, Y. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2886-2889 (1979).
- Patel, P. I. *et al. Somatic Cell and Molec. Genet.* (in the press).
- Yang, T. P., Hansen, S. K., Oishi, K. K., Ryder, O. A. & Hamkalo, B. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6593-6597 (1982).
- Higuchi, R. *et al. Gene* **15**, 177-186 (1981).
- Fuscoc, J. C., Fenwick, R. G., Ledbetter, D. H. & Caskey, C. T. *Molec. cell. Biol.* **3**, 1086-1096 (1983).

Fertilization increases the polyphosphoinositide content of sea urchin eggs

Paul R. Turner, Michael P. Sheetz & Laurinda A. Jaffe

Department of Physiology, University of Connecticut Health Center, Farmington, Connecticut 06032, USA

Fertilization of the sea urchin egg stimulates a wave of exocytosis of cortical vesicles^{1,2}, but the mechanism by which fertilization regulates this secretion is not fully understood. We describe here experiments which suggest that polyphosphoinositide metabolism could be a factor in this regulation. We find that the cortical vesicle exocytosis in the egg of *Strongylocentrotus purpuratus* is preceded by a 40% increase in its content of triphosphoinositide (TPI) and a 22% increase of diphosphoinositide (DPI).

Following addition of a concentrated sperm suspension to a sample of sea urchin eggs, fertilization occurs within 3 s (ref. 3). About 30-40 s later, vesicles begin to fuse with the egg plasma membrane, and secretion continues for ~20-60 s (*S. purpuratus* at 15-16°C; refs 4, 5). The exocytosis causes an envelope to form around the egg which provides a block to polyspermy⁶. Concomitant with the vesicle fusion and detectable about 26 s after fertilization (sea urchin *Arbacia punctulata* at 21°C; ref. 7), there is a rise in intracellular free calcium⁸. Although this rise is required for secretion to occur^{9,10}, other regulatory factors might also be necessary, for the release of the calcium, or to act synergistically with the calcium. Exocytosis depends on ATP¹¹ and might involve calmodulin¹²⁻¹⁴.

Our reasons for thinking that polyphosphoinositide metabolism might be involved in the regulation of cortical vesicle secretion were the following: (1) secretion in other cells is accompanied by changes in polyphosphoinositide metabolism¹⁵⁻¹⁸; (2) 25% of the phospholipids in the sea urchin egg plasma membrane are phospho- or polyphosphoinositides (about 3×10^{-15} moles per plasma membrane)¹⁹; (3) by 1 h post-fertilization, eggs loaded with ³H-inositol shown an increase in the proportion of label in the polyphosphoinositides compared with phosphoinositide²⁰. We therefore examined whether changes in the amounts of TPI and DPI occur rapidly enough to be important in the regulation of cortical vesicle exocytosis.

In initial experiments, samples of eggs labelled with ³²P as described in Fig. 1 legend, were spotted directly onto plates for TLC, using CHCl₃/CH₃OH/4M NH₄OH, 9:7:2 as the solvent²⁰. The positions of the lipids were identified by

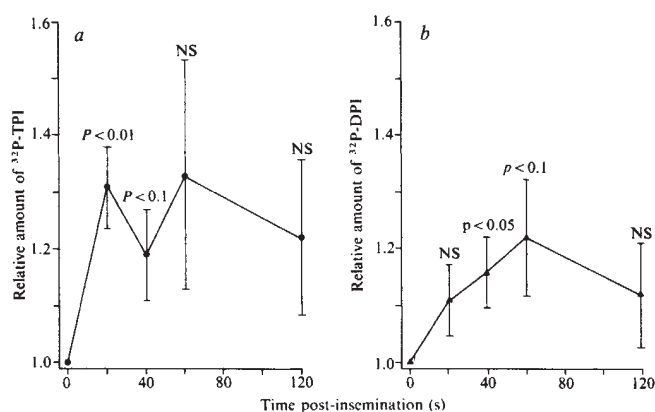


Fig. 1 Changes in the amount of ³²P in *a*, the TPI fraction and *b*, the DPI fraction, at various times after fertilization. Data are normalized relative to the amount of label in the unfertilized eggs (*t* = 0 s) in each of six experiments. Bars indicate s.e.m.; *P* values were determined by Student's *t*-test; NS, not significant.

Methods: Eggs were obtained by injecting animals with 0.5 M KCl. They were dejelled in normal seawater, pH 5.0 for 3 min. The egg suspension was adjusted to 5% by volume. Two ml of eggs were preincubated for 2 h in seawater containing ³²PO₄ (~0.25 mCi ml⁻¹; NEN); complete saturation of the ³²P pool of the unfertilized egg requires an incubation of longer than 28 h, at which time the eggs begin to disintegrate. Uptake is linear for at least 10 h (unpublished observations). Eggs were inseminated in the ³²PO₄-containing seawater by the addition of enough sperm to make a 0.1% sperm suspension in the final volume. The temperature during insemination was 15°C. 20-μl samples were taken before and at various times post-insemination, and were added to 5 vols of chloroform-methanol, 1:2. Then 1.4 vols of 1.2 M HCl-chloroform (1:1) were added, and the mixture was vortexed and spun for 2 min. The aqueous layer was removed and replaced with 1M HCl-methanol (1:1). After two additional washes, the chloroform layer was neutralized by adding an equal volume of 200 mM ammonium acetate in methanol. Samples were kept on dry ice. The polyphosphoinositides were purified from this total lipid extract by passing it over small columns containing neomycin bound to glass beads (100 μl of beads)²⁷. First, lipids other than the polyphosphoinositides were eluted with 20 column-volumes of 150 mM ammonium acetate, in chloroform-MeOH-H₂O, 3:6:1. DPI was then eluted by adding 10 column-volumes of 600 mM ammonium acetate in chloroform-MeOH-H₂O, 3:6:1. Finally, TPI was eluted by the addition of 20 column-volumes of chloroform-MeOH-14.8 M NH₄OH, 3:6:1. The DPI and TPI fractions were counted in a liquid scintillation counter (Czerenkoy).

autoradiography, and the spots were scraped and counted. A 20-μl sample of a 5% egg suspension contained about 4,000 eggs and about 2,000 c.p.m. in the TPI fraction. We found that, as previously reported for *A. punctulata*²⁰, most of the label taken up into lipids of unfertilized eggs of *S. purpuratus* is in TPI (44 ± 5%) and DPI (26 ± 3%), compared with 10 ± 6% for phosphatidylinositol and phosphatidic acid, and 20 ± 1% for other lipids; all values means ± s.e.m. for six experiments.

Next we measured the incorporation of ³²P into TPI and DPI before, and at various times after fertilization, as described in Fig. 1 legend. 50 ± 4% of the TPI and 53 ± 1% of the DPI counts (s.e.m., *n* = 4) were recovered after extraction and column chromatography. To confirm that the column procedure was effectively separating TPI and DPI, fractions eluted from the column were examined with TLC as described above. TPI and DPI were found only in the expected fractions and were free from contamination. At 20 s post-insemination, the TPI fraction contains 31% more label than the unfertilized fraction (*P* < 0.01). DPI showed a smaller increase in ³²P incorporation. Increases in labelling are unlikely to be due to an increased uptake of ³²P by the fertilized egg, because after fertilization there is a delay of about 5-15 min before the rate of phosphate uptake increases²¹.

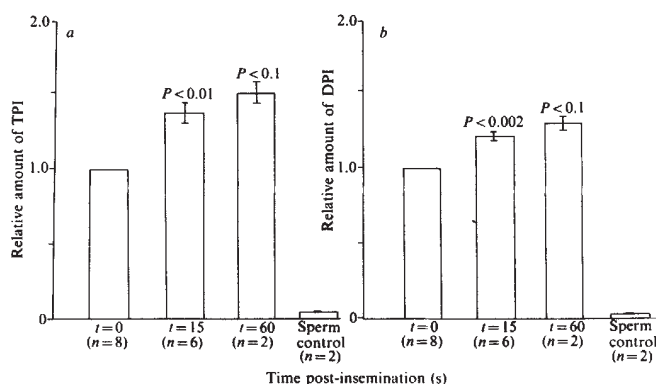


Fig. 2 Changes in the chemical amounts of *a*, TPI and *b*, DPI at various times after fertilization. Data are normalized relative to the amounts of TPI or DPI in the unfertilized eggs ($t = 0$ s) in each experiment. Bars indicate s.e.m., n the number of experiments. P values were determined by Student's t -test; NS, not significant. See text for explanation of sperm control experiments.

Methods: DPI and TPI were purified as described for Fig. 1. Ten ml of a 5% egg suspension was used for each sample. Columns were run at 0 °C. Fractions were dried and the amount of inorganic phosphate in each was determined²⁸.

Although our results are difficult to interpret without a knowledge of the specific activity of the ATP pool, they did encourage us to assay chemically the amounts of DPI and TPI present in samples of sea urchin eggs taken before and after fertilization (Fig. 2). Using column chromatography as described for Fig. 1, and analysing the phosphate content of the eluted fractions (see Fig. 2 legend), we found that each unfertilized egg contains $2.9 \pm 0.1 \times 10^{-15}$ (s.e.m., $n = 8$) moles of TPI and $6.9 \pm 0.7 \times 10^{-15}$ (s.e.m., $n = 8$) moles of DPI. Increases in amounts of both TPI and DPI are seen 15 s after fertilization, before vesicle exocytosis. At 15 s post-insemination, the amount of TPI is 40% greater than in unfertilized eggs ($P < 0.01$); at 60 s, the increase is 52%. DPI is increased by 22% at 15 s ($P < 0.002$) and by 31% at 60 s.

To exclude the possibility that the sperm used in the inseminations was providing the increased amounts of TPI and DPI seen following fertilization, we extracted these lipids from an amount of sperm equivalent to that used in the fertilization experiments. (The sperm had been exposed to solubilized egg jelly to induce the acrosome reaction—acrosomal processes were seen in 70% of the cells.) The sperm contain both DPI and TPI (see Fig. 2), but the amounts constitute only 9–15% of the changes seen at fertilization.

We conclude that there is an increase in the content of both TPI and DPI following fertilization, but before the onset of vesicle fusion, in the eggs of the sea urchin *S. purpuratus*. Available data from another species of sea urchin⁸ indicate that our observed increases in TPI and DPI also precede the rise in intracellular free calcium. These changes in TPI and DPI metabolism might lead to the release of calcium from intracellular stores. If the net increases we report indicate an increased rate of TPI turnover, the egg might be generating inositol-1,4, 5-trisphosphate, a substance known to release calcium from intracellular stores in pancreatic acinar cells²². Alternatively, TPI and DPI might contribute to the stimulation of secretion by a direct action occurring in parallel with the rise in free calcium, but not causing it. The presence of negatively charged lipids, and increases in membrane fluidity, both increase the incidence of vesicle fusion with cells²³. The polyphosphoinositides are highly charged at normal pH—TPI has a net charge of -5 , and DPI one of -3 (ref. 24)—which may promote fusion by increasing binding of calcium to the membranes. Moreover, both TPI, and to a lesser extent DPI, increase glycoprotein lateral mobility in the red blood cell membrane²⁵. These characteristics of the polyphosphoinositides might be important in initiating cortical vesicle exocytosis.

The increase in polyphosphoinositide content of the sea urchin egg might also be significant with respect to other changes in the egg's metabolism after fertilization²⁶. The increase in polyphosphoinositide content that we report precedes all previously reported events, except for the membrane depolarization, and is therefore the earliest known biochemical event following fertilization of the sea urchin egg.

We thank P. Cameron for help with the column chromatography. This work was supported by NIH grants HD-14939 (L.A.J.) and HL-23359 (M.P.S.). M.P.S. is an Established Investigator of the American Heart Association.

Received 30 December 1983; accepted 31 May 1984.

- Endo, Y. *Expl Cell Res.* **25**, 383–397 (1961).
- Chandler, D. E. & Heuser, J. J. *Cell Biol.* **83**, 91–108 (1979).
- Rothschild, L. & Swann, M. M. J. *exp. Biol.* **29**, 469–483 (1952).
- Paul, M. & Epel, D. *Expl Cell Res.* **65**, 281–288 (1971).
- Jaffe, L. A., Hagiwara, S. & Kado, R. T. *Dev Biol.* **67**, 243–248 (1978).
- Schuel, H. *Gamete Res.* **1**, 299–382 (1978).
- Eisen, A. thesis, Univ. Pennsylvania (1982).
- Steinhardt, R., Zucker, R. & Schatten, G. *Dev Biol.* **58**, 185–196 (1977).
- Zucker, R. S. & Steinhardt, R. A. *Biochim. biophys. Acta* **541**, 459–466 (1978).
- Hamaguchi, Y. & Hiramoto, Y. *Expl Cell Res.* **134**, 171–179 (1981).
- Baker, P. F. & Whitaker, M. J. *Nature* **276**, 513–515 (1978).
- Steinhardt, R. A. & Alderton, J. M. *Nature* **295**, 154–155 (1982).
- Picard, A. & Dorée, M. *Dev Growth Different.* **24**, 155–162 (1982).
- Whitaker, M. J. & Baker, P. F. *Proc. R. Soc. B218*, 397–413 (1983).
- Billah, M. M. & Lapetina, E. G. *J. biol. Chem.* **257**, 12705–12708 (1982).
- Weiss, S. J., McKinney, J. S. & Putney, J. W. *Biochem. J.* **206**, 555–560 (1982).
- Agranoff, B. W., Murthy, P. & Seguin, E. B. *J. biol. Chem.* **258**, 2076–2078 (1983).
- Perret, B. P., Plantavid, M., Chap, H. & Douste-Blazy, L. *Biochem. biophys. Res. Commun.* **110**, 660–667 (1983).
- Kinsey, W. H., Decker, G. L. & Lennarz, W. J. *J. Cell Biol.* **87**, 248–254 (1980).
- Schmell, E. & Lennarz, W. J. *Biochemistry* **13**, 4114–4121 (1974).
- Chambers, E. L. & Whiteley, A. H. J. *cell. Physiol.* **68**, 289–308 (1966).
- Streb, H., Irvine, R. F., Berridge, M. J. & Schulz, I. *Nature* **306**, 67–69 (1983).
- Papahadjopoulos, D., Poste, G. & Schaeffer, B. E. *Biochim. biophys. Acta* **323**, 23–42 (1973).
- Downes, P. & Michell, R. H. *Cell Calcium* **3**, 467–502 (1982).
- Scheetz, M. P., Febroriello, P. & Koppel, D. E. *Nature* **296**, 91–93 (1982).
- Epel, D. *Curr. Topics dev. Biol.* **12**, 185–246 (1978).
- Schacht, J. J. *Lipid Res.* **19**, 1063–1067 (1978).
- Ames, B. N. *Meth. Enzym.* **8**, 115–118 (1966).

Regulated expression of an introduced MHC H-2K^{bm1} gene in murine embryonal carcinoma cells

Arnon Rosenthal*†, Stephanie Wright*, Howard Cedar†, Richard Flavell*‡ & Frank Grosveld*

* Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

† Department of Molecular Biology, The Hebrew University, Hadassah Medical School, Jerusalem, Israel

‡ Biogen Research Corporation, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA

The transplantation antigens H-2K, H-2D and H-2L are developmentally regulated^{1–3}, highly polymorphic⁴ cell surface proteins encoded by the major histocompatibility gene complex (MHC)^{5–7}. First detectable on the early embryo^{2,3}, they are subsequently expressed on most somatic cells of the adult mouse in association with the protein β_2 -microglobulin (β_2 M; ref. 5). Cultured F9 embryonal carcinoma (EC) cells can be induced to differentiate along alternative pathways to form either parietal⁸ or visceral⁹ extra-embryonic endoderm, each concomitant with a change in morphology and pattern of gene expression. Previous reports have demonstrated an increased level of transplantation antigens in differentiated F9 EC cells^{10–14}, but the cell types expressing them were not defined. Here we show that the level of MHC H-2K^b and β_2 M transcripts is increased during both pathways of this differentiation. Expression of a foreign MHC H-2K^{bm1} gene was found to be regulated in a similar manner when the gene was introduced into EC cells. In contrast, an introduced rabbit β -globin gene was not regulated but expressed constitutively.

F9 EC cells were differentiated into parietal endoderm by culturing as a monolayer with 10^{-7} M retinoic acid for 14 days⁸, or with 10^{-7} M retinoic acid plus 10^{-4} M N^6, O^2 -dibutyryl-adenosine 3',5'-cyclic monophosphate (dbcAMP) plus 10^{-4} μ M 3-isobutylmethylxanthine for 5 days, conditions in which 90% of the cells are differentiated^{15,16}. For differentiation to visceral endoderm, cells were grown for 6 days as small floating aggregates in bacteriological Petri dishes with 5×10^{-8} M retinoic acid⁹. MHC H-2K^b and β_2 M mRNA transcripts were quantitated by S₁ nuclease analysis using an exon-intron and a cDNA probe respectively, which can detect ~10 pg of specific mRNA after densitometric scanning.

Using a 161-nucleotide *Hpa*II fragment spanning part of the β_2 M cDNA and adjacent plasmid sequences as a probe for S₁ nuclease protection, no β_2 M transcripts were detected in undifferentiated F9 cells (Fig. 1). Differentiation to parietal or visceral endoderm resulted in at least a 20-fold increase in the level of β_2 M transcripts, as demonstrated by the presence of 110- and 80-nucleotide fragments resistant to digestion by S₁ nuclease (Fig. 1). This is consistent with the high level of this transcript observed in visceral and parietal endoderm isolated from the extra-embryonic tissues of a 13.5-day-old mouse. The shorter than full-length 80-nucleotide fragment may be due to either a small difference between the cDNA probe and the transcript or cleavage by S₁ nuclease of an A+T-rich region located 85 base pairs (bp) from the labelled *Hpa*II site¹⁷.

S₁ nuclease analysis using a 600-nucleotide *Ava*II exon III-intron probe (Fig. 2b) was used to determine the level of H-2K^b mRNA. As judged by the presence of a 230-nucleotide protected fragment, this transcript is produced in very small quantities in undifferentiated F9 cells. After differentiation either to visceral endoderm using retinoic acid or to parietal endoderm using retinoic acid alone or in combination with dbcAMP (Fig. 2b), the level of H-2K mRNA increased 5–10-fold as determined by densitometer scanning.

Analysis of RNA samples from other undifferentiated and differentiated F9 cells (given by B. Hogan)^{9,15,16}, showed levels of H-2K^b mRNAs identical to ours (not shown). Low levels of H-2K^b mRNA in undifferentiated EC cells have been reported previously^{18,19} and resemble the results described for laminin. Laminin mRNA is also transcribed at low levels in undifferentiated F9 cells and its transcription is increased following differentiation¹⁵. The fact that β_2 M RNA is not detectable in undifferentiated F9 cells may indicate that the H-2K and β_2 M genes are regulated differently. Note that despite the low levels of H-2K^b RNA synthesis, H-2K^b antigen might not be detected^{10,20} on the cell surface in the absence of β_2 M.

Experiments using cell fusion or chromosome transfer between somatic cells and EC cells have shown that the somatic H-2K gene is either repressed²¹ or remains active in the hybrid cells^{19,22,23}. As these experiments were inconclusive, we have used DNA-mediated gene transfer to study the control of H-2K expression.

A 10-kilobase (kb) fragment containing a variant H-2K^b gene called H-2K^{bm1} was introduced into F9 cells²⁴. H-2K^{bm1} was covalently linked to Simian virus 40 (SV40) DNA and to the transposon Tn5-derived aminoglycosyl phosphotransferase II (AGPT) as a selective marker (Fig. 2a). Using the calcium phosphate transformation method²⁵, 1–10 clones per 5×10^5 cells per μ g DNA were obtained. Southern blotting showed that 12 transformants contained 1–15 copies of the introduced DNA per cell (data not shown).

Taking advantage of the fact that the introduced H-2K^{bm1} gene and the endogenous H-2K^b gene differ in the third exon sequence²⁴, we prepared a 600-bp H-2K^b *Ava*II probe from this region which allowed us to distinguish between the transcripts from the two genes by S₁ digestion analysis (Fig. 2b). Because partial S₁ nuclease digestion of the H-2K^b *Ava*II probe protected by H-2K^{bm1} mRNA yields the same fragment as that protected by the H-2K^b transcript, we confirmed the results using the corresponding probe isolated from the H-2K^{bm1} gene (not shown).

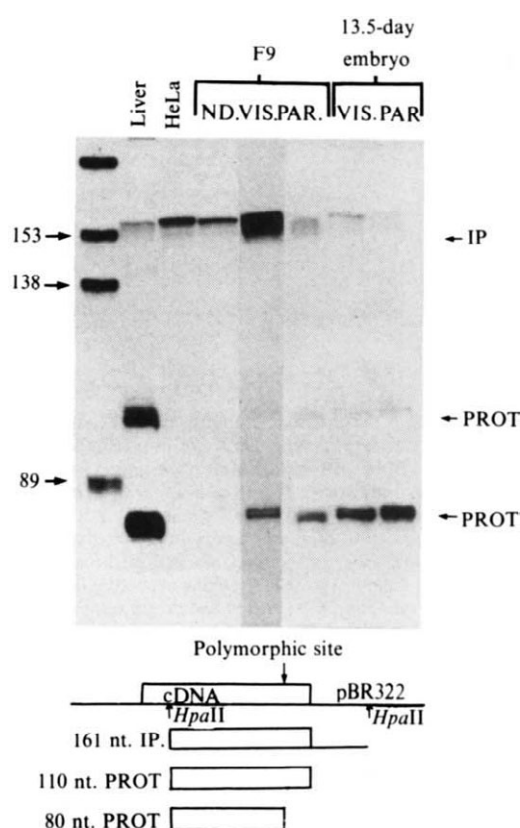


Fig. 1 S₁ nuclease analysis of β_2 M mRNA in F9 EC cells before (ND) and after differentiation to parietal (PAR) or visceral (VIS) endoderm. Control mRNA was obtained from visceral and parietal extra-embryonic endoderm of a 13.5-day-old mouse embryo, and from adult mouse liver. Lane 1 contains a radioactivity labelled Φ X $\times$ *Rsa* marker.

Methods: A cDNA probe for β_2 M (IP) (illustrated at bottom of figure) was end-labelled by reverse transcription, strand-separated and hybridized with total RNA at 37 °C for 12 h in 10 μ l of 80% formamide, 40 mM PIPES pH 6.4, 1 mM EDTA, 400 mM NaCl. The mixture was digested with 3,000 U of S₁ nuclease (Boehringer) in 300 μ l of 300 mM NaAc pH 4.8, 200 mM NaCl, 2 mM ZnSO₄ at 20 °C for 2 h. S₁-protected DNA (Prot.) was ethanol-precipitated, denatured and electrophoresed on a 7M urea/7% acrylamide gel. F9 lanes: ND, 15 μ g of undifferentiated F9 cell RNA (overexposed fivefold); PAR, 15 μ g of parietal endoderm F9 cell RNA. Embryo lanes: 5 μ g of visceral or parietal endoderm (from B. Hogan).

Out of eight clones that contained intact copies of the H-2K^{bm1} introduced DNA, seven clones showed a 6–15-fold increase in the level of H-2K^{bm1} mRNA on differentiation to parietal endoderm (Fig. 2b, clones 2–5). Only one clone expressed the gene in a constitutive manner (not shown). Four clones were shown to contain an incomplete or rearranged copy of the introduced DNA and on differentiation, only the endogenous H-2K^b transcripts could be detected (Fig. 2b, clone 1). Note that the level of the endogenous H-2K^b mRNA is 5–10-fold lower than the level found in liver and spleen of H-2K^b mice. In contrast, the exogenous H-2K^{bm1} gene is expressed at higher levels. This effect may be due either to a higher copy number (see legend to Fig. 2b) or to selective integration sites.

The increased levels of H-2K mRNA observed on differentiation of EC cells could be a result of transcriptional activation or specific stabilization of H-2K mRNA; we cannot yet distinguish between these possibilities.

All of the above S₁ analyses were done using a probe which spans one of the exon-intron junctions within the gene. To

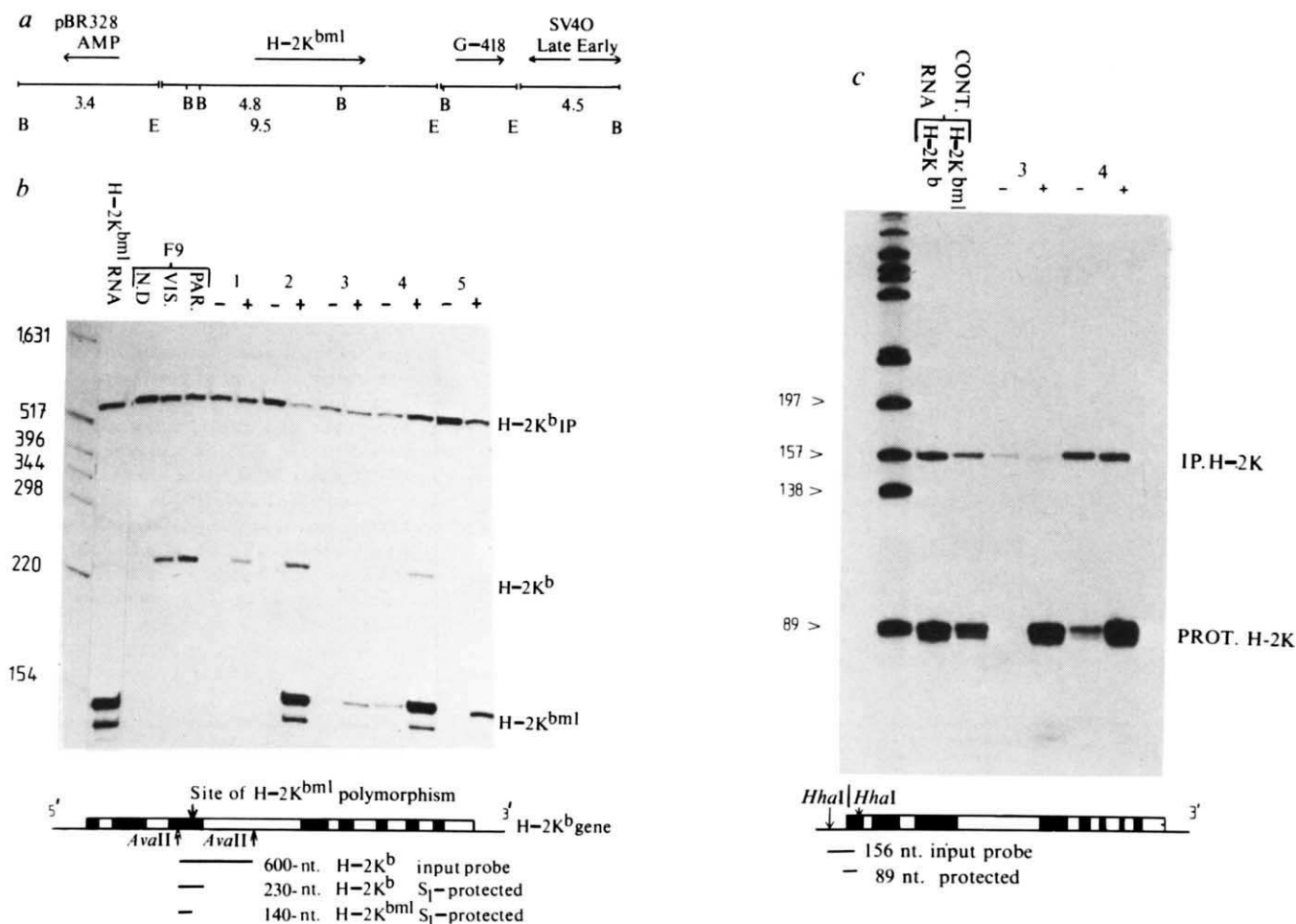


Fig. 2 *a*, Structure of the plasmid introduced into F9 cells. Transcripts are illustrated by arrows in the 5' to 3' direction. E and B, *EcoRI* and *BamHI* sites, respectively. *b*, S₁ nuclease analysis of H-2K^b and H-2K^{bml} RNA from untransformed (F9) and from H-2K^{bml}-transformed (1-5 EC cells) before (-) and after (+) differentiation. Lane labelled H-2K^{bml} RNA, 10 µg of RNA from a transient expression of the H-2K^{bml} gene in HeLa cells³¹. F9 lanes: 10 µg of RNA as in Fig. 1. Lanes 1-5, 10 µg of transformed F9 cell RNA containing, respectively, 1-3, 5-8, 3-5, 10-12 and 3-5 integrated copies of the H-2K^{bml} gene. *c*, S₁ nuclease analysis of the 5' end of H-2K mRNA in two F9 transformants, before (-) and after (+) differentiation.

Methods: *b*, A 600-nucleotide intron-exon probe that allowed a distinction to be made between H-2K^{bml} and H-2K^b transcripts (see *a*) was end-labelled with reverse transcriptase, strand-separated and hybridized as in Fig. 1, except that S₁ digestion was for 1 h at 20 °C followed by 1 h at 40 °C. Lane 4 contains a pBR322 × *Hinf* marker. *c*, The 5' H-2K^b probe was end-labelled with T4 polynucleotide kinase, strand-separated and hybridized as described in Fig. 1.

determine whether these transcripts were properly initiated, we used a probe that covered the 5' end of the H-2K^b and H-2K^{bml} genes, which are identical in this region. Although this probe does not distinguish between the mRNA produced from the endogenous and exogenous genes, it showed only the correct 5' end expected for H-2K^b and H-2K^{bml} mRNAs, therefore we conclude that transcription of the exogenous gene is correctly initiated (Fig. 2c).

In another set of transfection experiments, the rabbit β-globin gene, covalently linked to the same SV40 vector, was introduced into F9 EC cells. S₁ nuclease analysis of four different transformants showed a low level of β-globin mRNA which did not increase following differentiation, while the endogenous H-2K^b mRNA level in the same transformants increased (Fig. 3).

These results suggest that the DNA sequences involved in this induction process must be present on the 10-kb fragment containing the H-2K^{bml} gene used in this experiment and that these sequences are specific for the H-2K gene. It is highly

unlikely that the control sequences would be tissue-specific enhancer elements, as first observed for immunoglobulin genes²⁶⁻²⁸; instead, they may control the activation of the H-2K gene at a certain time in the development programme and its subsequent expression in all somatic cells. Induction of this sort may be mediated either by the release of gene suppression or by gene activation and our data are compatible with both of these models.

The fact that visceral-like endoderm and parietal-like endoderm F9 cells, which differ morphologically and functionally, both express H-2K and β₂M may imply that the F9 EC cells originate from embryonic cells predetermined²⁹ to express transplantation antigens. Thus, although F9 EC cells can be differentiated along alternative pathways^{8,9,30}, with respect to the expression of MHC genes they may not be different from other, already committed cell systems where increased gene expression can be induced chemically.

This work was supported by the BMRC (UK). A.R. was

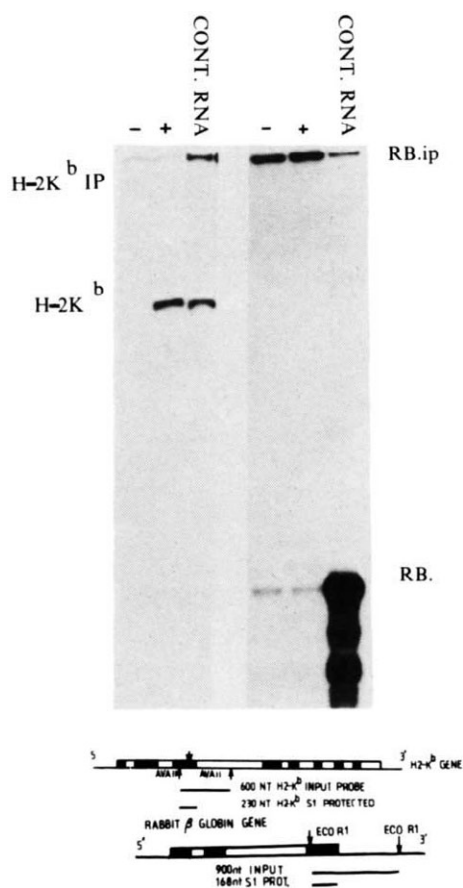


Fig. 3 S₁ nuclease analysis of H-2K^b and rabbit β -globin mRNA before (–) and after (+) differentiation. Control mRNA for H-2K^b was obtained from nontransformed F9 cells after differentiation to parietal endoderm. Control mRNA for rabbit β -globin was obtained from rabbit bone marrow. The 3' rabbit β -globin probe (RBip) was end-labelled using reverse transcriptase and hybridized as described in Fig. 1 legend.

supported by The Michael and Anna Wix Trust through the Friends of the Hebrew University. We thank D. Barlow and B. Hogan for gifts of RNA and for technical advice, and R. Axel for his hospitality and assistance in the initial transformations.

Received 14 February; accepted 21 May 1984.

- Strickland, S., Smith, K. K. & Maratti, K. K. *Cell* **21**, 347–355 (1980).
- Ostrand-Rosenberg, S. *Curr. Topics dev. Biol.* **14**, 147–168 (1980).
- Webb, C. G. *et al. J. exp. Med.* **146**, 923–932 (1977).
- Klein, J. *Science* **203**, 516–521 (1977).
- Ploegh, H. L., Orr, H. T. & Strominger, J. L. *Cell* **24**, 287–299 (1981).
- Hood, L., Steinmetz, M. & Goodnow, R. *Cell* **28**, 685–687 (1982).
- Klein, J. *Biology of the Mouse Major Histocompatibility Complex* (Springer, Berlin, 1975).
- Strickland, S. & Mahdavi, V. *Cell* **15**, 393–403 (1980).
- Hogan, B. L. & Taylor, A. *Nature* **291**, 235–237 (1981).
- Nicolas, J. F. *et al. Ann. Microbiol.* **A126**, 3–22 (1975).
- Croce, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 5754–5748 (1981).
- Knowles, B. B. *et al. Nature* **288**, 615–617 (1980).
- Jacob, F. *Immunol. Rev.* **33**, 3–32 (1977).
- Tanaka, K., Appella, E. & Jay, G. *Cell* **35**, 457–465 (1983).
- Cooper, A. R., Taylor, A. & Hogan, B. L. M. *Dev. Biol.* **99**, 510–516 (1983).
- Kurkinen, M. *et al. Nucleic Acids Res.* **11**, 6200–6209 (1983).
- Parnes, J. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2253–2257 (1981).
- Sherman, M. I. & Miller, R. A. *Dev. Biol.* **63**, 27–34 (1978).
- Benham, P. J., Quintero, M. A. & Goodfellow, P. M. *EMBO J.* **2**, 1963–1968 (1983).
- Artzt, K. & Jacob, F. *Transplantation* **17**, 632–625 (1975).
- Forejt, J. *Cold Spring Harb. Symp. quant. Biol.* **48**, 541–552 (1983).
- Andrews, D. W. & Goodfellow, P. N. *Somat. Cell Genet.* **6**, 271–284 (1980).
- Correani, A. & Croce, C. M. *J. cell. Physiol.* **105**, 73–99 (1980).
- Weiss, E. H. *et al. Nature* **301**, 671–674 (1983).
- Wigler, G. R. *et al. Cell* **16**, 777–785 (1979).
- Queen, C. & Baltimore, D. *Cell* **33**, 741–748 (1983).
- Gillies, S. D. *et al. Cell* **33**, 719–728 (1983).
- Banerji, J. *et al. Cell* **33**, 729–740 (1983).
- Murphy, D. *et al. Cell* **35**, 865–871 (1983).
- Liesi, P. *et al. Nature* **306**, 265–267 (1983).
- Grosfeld, G. C. *et al. Nature* **295**, 120–126 (1982).

Antibodies to a scrapie prion protein

Paul E. Bendheim*, Ronald A. Barry*,
Stephen J. DeArmond†, Daniel P. Stites§
& Stanley B. Prusiner*†

Departments of * Neurology, † Biochemistry and Biophysics,
‡ Pathology, and § Laboratory Medicine, University of California,
San Francisco, California 94143, USA

Scrapie is a slow infection of the nervous system which progresses in the absence of any apparent immune response^{1–8}. The recent development of a large-scale purification protocol for scrapie prions⁹ made it possible to obtain substantial quantities of electrophoretically purified prion protein (PrP 27–30) and we report here on the successful production of a rabbit antiserum to PrP 27–30. The antiserum reacted with PrP 27–30 and several lower molecular weight proteins as shown by Western blots; it did not react with protein preparations from uninfected brains. Discrete structures in the subependymal region of scrapie-infected hamster brains were stained immunocytochemically. These same structures also stained with Congo red dye and showed green birefringence with polarized light, a characteristic of purified prion rods⁹. This staining pattern suggests that they are amyloid plaques.

Although the immune system remains competent in both natural and experimental scrapie, prions fail to evoke humoral or cell-mediated responses^{1–8}. The host is afebrile, leukocyte counts in the cerebrospinal fluid and blood are normal, and no inflammatory response is seen. Nevertheless, the animal dies of an overwhelming central nervous system slow infection. Progress in the purification of the scrapie agent⁹ led to the identification of a protease-resistant protein which is a major component of the prion^{10,11}. This protein (PrP 27–30) aggregates into rod-like structures which have the ultrastructural^{12,13} and histochemical characteristics of amyloid¹².

Three rabbits were immunized with 10–20 μ g of PrP 27–30, coupled to keyhole limpet haemocyanin (KLH), and boosted twice with 10–20 μ g of PrP–KLH. These animals failed to produce antibodies to PrP 27–30. A fourth rabbit immunized with 80 μ g of PrP 27–30 without KLH and boosted with 130 μ g of PrP 27–30 after 25 days produced antibodies. For subsequent boosts in this rabbit, denatured scrapie prions purified by zonal rotor sucrose gradient sedimentation were used without further electrophoretic purification of PrP 27–30. We estimate that the third and fourth boosts contained 40 μ g and 100 μ g of PrP 27–30, respectively.

The sera were tested by a modified dot-blot immunoassay¹⁴. Preimmune serum (Fig. 1, row a) did not react with purified PrP 27–30, while only immune serum from the fourth rabbit reacted with purified PrP 27–30 (Fig. 1, rows b–e). Reactivity with PrP 27–30 was seen at a 1:1,000 dilution of the serum collected 25 days after the first immunization. Immune sera collected after subsequent boosts showed a progressive rise in titre. After three immunizations, the serum could be diluted 1:20,000 and a positive response was seen with purified PrP 27–30.

Immune serum failed to react with proteins purified from uninfected brains by the same protocol (Fig. 1, rows f, g). This antiserum reacted with both denatured and native prions purified from scrapie-infected hamster brains (Fig. 1, rows h, i). Preimmune serum at a 1:100 dilution gave a detectable response against the denatured and native prions (data not shown), but failed to react with electrophoretically purified PrP 27–30 (Fig. 1, row a). At 1:1,000 dilution, no reaction of the preimmune serum with denatured or native prions could be detected. With antiserum diluted 1:1,000, PrP 27–30 could be detected even after its concentration was reduced by a factor of 100 (Fig. 1, rows j, k).

Western blots^{15,16} were prepared to confirm that the antiserum was reacting specifically with PrP 27–30 (Fig. 2). PrP 27–30 and

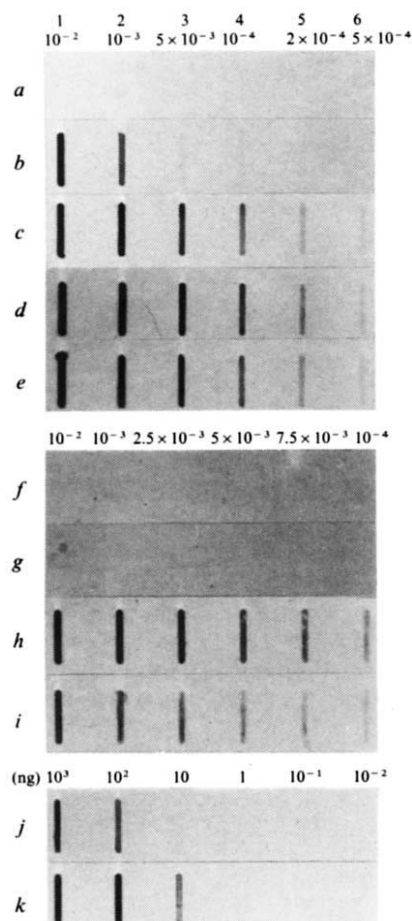


Fig. 1 Immunoreactivity of anti-PrP rabbit serum. The reactivity of sera diluted as shown, to 100 ng of electrophoretically purified PrP 27-30: preimmune serum (*a*), and after the first (*b*), second (*c*), third (*d*), and fourth (*e*) immunizations. The reactivity of antiserum after the fourth immunization, diluted as shown, was tested against 100 ng of normal brain protein; *f*, denatured and *g*, native, as well as 100 ng of purified prions; *h*, denatured and *i*, native. The reactivities of antisera after the first and fourth immunization are shown in rows *j* and *k*, respectively, diluted $\times 10^{-3}$, and tested against increasing dilutions of antigen as shown. **Methods:** A female New Zealand white rabbit was immunized with PrP 27-30 purified by preparative SDS-gel electrophoresis of sucrose gradient fractions obtained by sedimentation of the prions in either a vertical² or zonal rotor¹⁰. Prions from gradient fractions were digested with proteinase K (25 $\mu\text{g ml}^{-1}$) for 60 min at 37 °C, then denatured by boiling for 3 min in 2% (w:v) SDS and 5% (v:v) β -mercaptoethanol before electrophoresis in 15% polyacrylamide vertical slab gels²⁷. The PrP 27-30 was eluted from the gel, emulsified in complete Freund's adjuvant and injected bilaterally into surgically exposed popliteal lymph nodes²⁸. Additional immunogen was injected into several subcutaneous sites over the back of the rabbit. At 3-4-week intervals, the rabbit was boosted with PrP 27-30 in incomplete Freund's adjuvant by subcutaneous injections. The antiserum was evaluated using a modified dot-blot immunobinding assay¹⁴. Samples to be tested for immunoreactivity were suspended in 50 mM Tris-HCl pH 8.1, 200 mM NaCl and 0.04% NaN₃ (TBS) and applied in two 50- μl aliquots to a nitrocellulose filter mounted in a vacuum manifold apparatus (Minifold II, Schleicher & Schuell). After air drying under low vacuum and washing in TBS, unabsorbed sites were blocked with 10% heat-inactivated horse serum and 3% (w:v) bovine serum albumin in TBS. The primary antisera, at the indicated dilutions, and the second antibody, goat anti-rabbit IgG horseradish peroxidase conjugated IgG (Tago) diluted 1:1,000, were applied in 150- μl aliquots. Immunoreactivity was detected using 4-chloro-1-naphthol and hydrogen peroxide¹⁴.

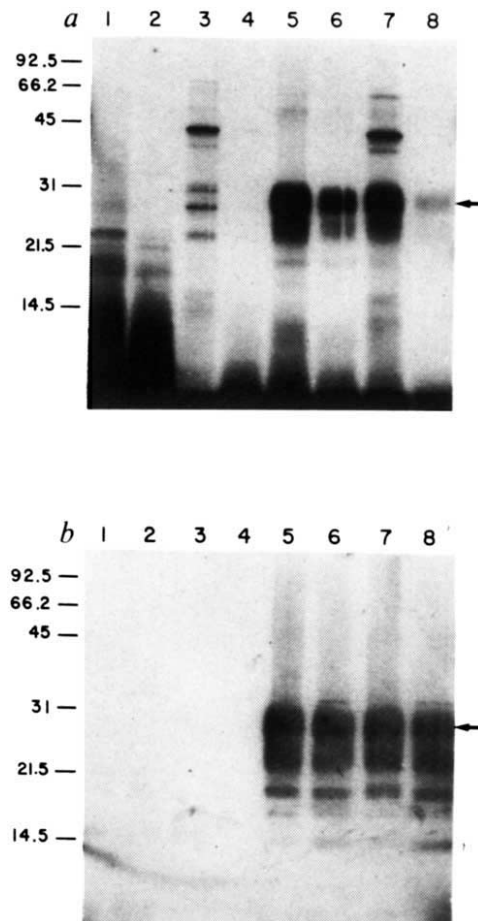


Fig. 2 Western blot of scrapie-infected and normal hamster brain preparations. *a*, Autoradiogram of proteins radioiodinated with the Bolton-Hunter reagent²⁹ and separated by electrophoresis. Lanes 1 and 3 are two different undigested preparations from normal hamster brains; lanes 2 and 4 are aliquots of the same samples digested with proteinase K. Lanes 5 and 7 are scrapie-infected samples; lanes 6 and 8 are aliquots of the same samples digested with proteinase K. *b*, Autoradiogram of immunoblot prepared using radioiodinated second antibodies. Samples were identical to those in *a* but unlabelled before electrophoresis in the polyacrylamide gel.

Methods: Samples of sucrose gradient fractions from two different scrapie-infected preparations and two different normal brain preparations were prepared in duplicate. Aliquots of radioiodinated and unlabelled fractions were incubated for 60 min at 30 °C in 10 mM Tris-HCl, pH 7.4, containing 0.2% (w:v) sodium dodecyl sarcosinate; half of the samples were digested with 25 $\mu\text{g ml}^{-1}$ of proteinase K. The digestion was terminated by the addition of an equal volume of 2 $\times$ concentrated SDS electrophoresis sample buffer and heating at 100 °C for 3 min. Electrophoresis was performed in 15% polyacrylamide gels according to the procedure of Laemmli²⁷. Proteins were electrophoretically transferred to nitrocellulose sheets^{15,16}. In *b*, after electrophoretic transfer to the nitrocellulose paper, the proteins were incubated for 1 h at room temperature in blocking buffer (100 mM Tris-HCl pH 7.4, 0.9% NaCl and 2 mg ml⁻¹ bovine serum albumin). The blotted proteins were incubated for 2 h at room temperature with a 1:1,000 dilution of the rabbit antiserum. Following three washes with blocking buffer, the second antibody (¹²⁵I-labelled goat anti-rabbit IgG, NE Nuclear) at a 1:5,000 dilution was added and incubated for 5 h at room temperature. The immunoblot was washed three times in blocking buffer, twice in double-distilled water and air dried. The autoradiogram was exposed for 4 days at room temperature. The molecular weights ($\times 10^{-3}$) of reference proteins are shown. Arrows show position of PrP 27-30.

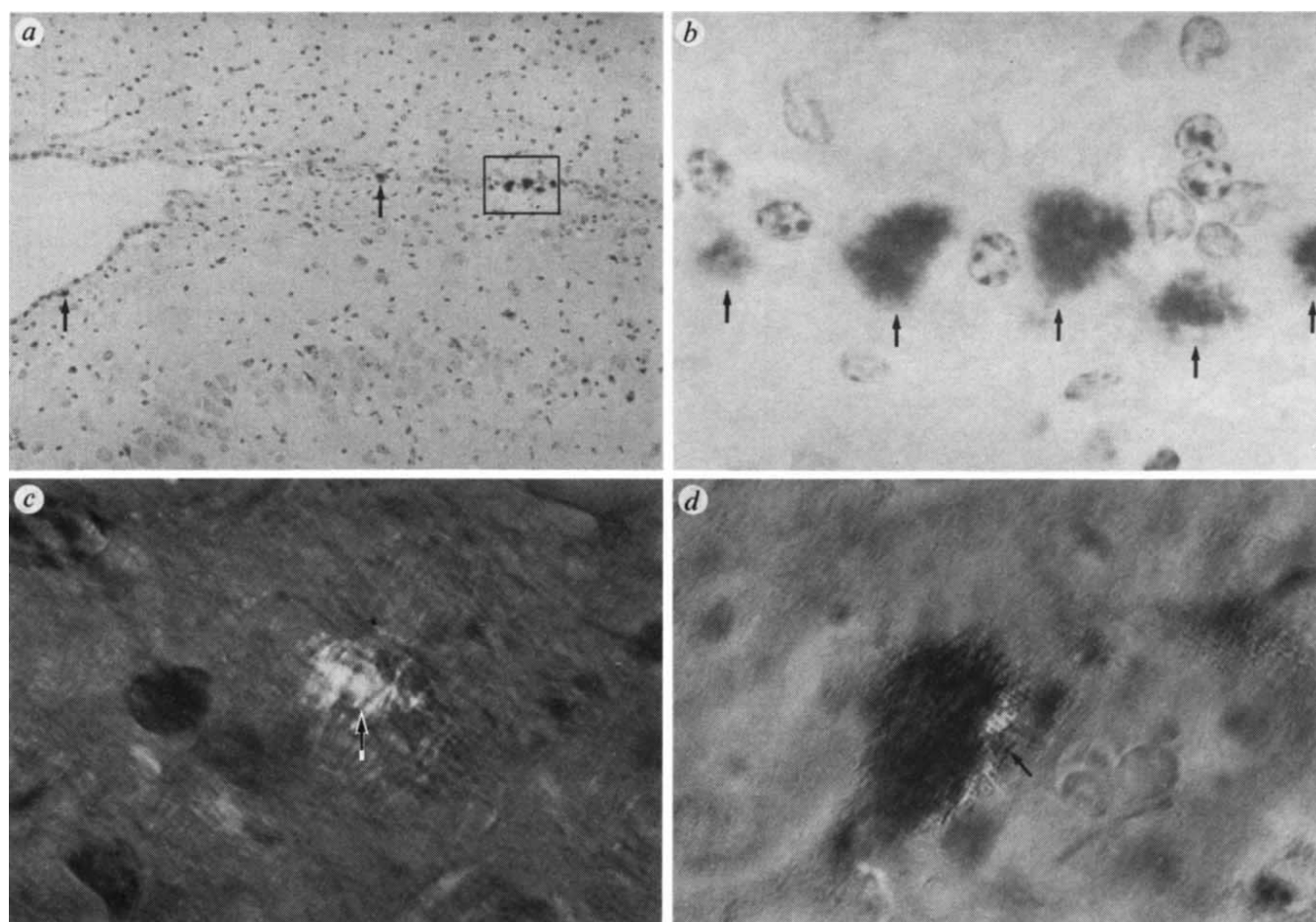


Fig. 3 Localization of PrP in scrapie-infected hamster brains by immunoperoxidase and Congo red histochemistry. *a*, PrP-positive structures (primary antiserum dilution 1:250) in the subependymal region adjacent to the hippocampal formation shown by arrows. Boxed region is that shown in *b*. Peroxidase-antiperoxidase stain with haematoxylin counterstain. $\times 70$. *b*, Same as *a* but $\times 700$. *c*, Congo red stained serial section of same region depicted in *a* and *b* showing green birefringence (arrow) with polarized light. $\times 1,000$. *d*, Peroxidase-antiperoxidase-stained serial section of the same region as depicted in *a* and *b* counterstained with Congo red, examined under polarized light. $\times 1,000$.

Methods: Hamsters were killed 65–70 days after intracerebral inoculation with 10^7 LD₅₀ units of hamster-adapted scrapie prions³⁰. Their brains were fixed in 10% formalin, embedded in paraffin and cut into 8- μ m sections. The peroxidase-antiperoxidase method¹⁷ was used to immunostain these sections. Diaminobenzidine-tetrahydrochloride was used to develop colour. Congo red staining was performed by the method of Puchtler *et al.*³¹. Photomicrographs were taken in a Leitz Orthoplan microscope.

several proteins of lower molecular weight were identified in blots from scrapie-infected hamster brains, but no proteins from uninfected hamster brains were immunolabelled. The pattern of immunoreactivity demonstrated a major protein of 27,500–32,000 molecular weight (MW) corresponding to PrP 27–30. The lower MW proteins—23,000–26,000 (23–26K), 19–20.5K, 17K, 14.5K—appear to share antigenic determinants with PrP 27–30; they either represent cleavage products of PrP 27–30 or other distinct prion-associated proteins. These lower MW proteins seem to exist in low concentrations based on autoradiograms of SDS-polyacrylamide gels (Fig. 2*a*, lanes 6, 8). Similar results were obtained with silver-stained polyacrylamide gels (data not shown).

PrP 27–30 and the other proteins identified by Western blots did not lose their antigenicity when exposed to limited digestion by proteinase K (Fig. 2*b*, lanes 6, 8). Additionally, proteinase K digestion of normal samples did not yield degradation products which were recognized by this antiserum (Fig. 2*b*, lanes 2, 4). Earlier studies had demonstrated that PrP 27–30 is relatively resistant to hydrolysis by proteinase K in comparison with other proteins found in infected hamster brains^{10,11}. Since proteinase K was used in the purification of PrP 27–30 it is not unexpected that antibodies were produced against components of this enzyme preparation. The immunoreactive protein of 33,000 MW (Fig. 2*b*, lanes 2, 4, 6, 8) represents an oligomer of proteinase K or a contaminant. Our results demonstrate that this antiserum

recognized PrP 27–30 and lower MW proteins which are relatively protease-resistant and not found in normal brain preparations.

The location of the scrapie prion proteins in three infected hamster brains was studied by peroxidase-antiperoxidase immunocytochemistry of formalin-fixed, paraffin-embedded tissue¹⁷. Positive staining was observed with primary antiserum dilutions ranging over 10^{-2} – 10^{-3} . In each animal, positive staining was highly localized to the subependymal region, especially adjacent to the hippocampus (Fig. 3*a*). The structures stained with PrP antiserum were discrete, measured up to 40 μ m in diameter and had a fluffy, cotton-like appearance (Fig. 3*b*). No other brain components were stained by this method and no staining was detected in three uninfected hamster brains. The preimmune rabbit serum stained neither the PrP-positive structures nor any other brain areas. When adjacent sections were stained with Congo red dye, these same structures exhibited a red colour by bright field microscopy and green birefringence with polarized light (Fig. 3*c*). Our results suggested that the structures staining with PrP antiserum were composed of paracrystalline arrays of prion proteins that exhibit the staining characteristics of amyloid plaques. To confirm these observations, sections stained with PrP antiserum were counterstained with Congo red dye and examined by polarization microscopy. As shown in Fig. 3*d*, the immunoperoxidase-stained structure also exhibited green birefringence.

Staining with Congo red dye and green birefringence of the PrP-positive structures *in situ* is consistent with the earlier observations that purified scrapie prions aggregate to form amyloid-like birefringent rods¹². Amyloid plaques have been found in both natural and experimental scrapie, but they are not an obligatory feature of the disease¹⁸. Certain inbred strains of mice inoculated with specific isolates of the scrapie agent have been shown to develop numerous amyloid plaques¹⁹. Presumably the structures reported here represent early amyloid plaques. Preliminary studies using immunoelectron microscopy with second antibody-colloidal gold have shown that both this antiserum and another rabbit antiserum raised against native prion rods react with PrP 27–30 in prion rods (R.A.B., M. P. McKinley, P.E.B., M. B. Braunfeld and S.B.P., in preparation).

The significance of the anatomical localization of PrP-positive structures within the subependymal areas and the hippocampus of scrapie-infected hamster brains is unclear. The hippocampus is the site of the first detectable light microscopic changes during scrapie infection in the hamster²⁰. The immunostaining of scrapie-infected hamster brains seems to reflect the high concentrations of PrP molecules deposited within the paraventricular regions. Perhaps these areas are selectively vulnerable to the accumulation of amyloid-like molecules. Earlier studies showed that scrapie infectivity is evenly distributed throughout the entire hamster brain²⁰; it is possible that prion proteins pass through the extracellular space of the brain and accumulate in these paraventricular areas. It is of interest that amyloid plaques in mice with experimental Creutzfeldt-Jakob disease were found mainly in the paraventricular cerebral white matter²¹ and that amyloid accumulations in the brains of normal aged humans are predominantly found in the subependymal areas of the brain as well as in the pia and septum pellucidum²².

The antiserum to hamster scrapie PrP 27–30 also reacts with proteins purified from the brains of mice with experimental Creutzfeldt-Jakob disease (P.E.B., J. Bockman, M. P. McKinley, D. T. Kingsbury and S.B.P., in preparation). The Creutzfeldt-Jakob disease associated proteins were purified by the same protocol used to isolate scrapie prions (D. T. Kingsbury, D. Smeltzer and J. Bockman, in preparation). Preliminary studies have shown that immunoreactive proteins of similar MW are present in purified fractions, prepared from the brains of two Creutzfeldt-Jakob disease patients (J. Bockman, P.E.B., S.B.P. and D. T. Kingsbury, unpublished observations). That the prions associated with these two diseases share antigenic determinants is consistent with earlier studies which demonstrated that the biological and physical properties of these slow infectious agents are similar^{23–25}.

The production of antisera that react with proteins from both these diseases should provide the basis for an immunoassay as well as lead to an affinity purification procedure. The availability of specific antibodies should aid both in structural studies of the prion and in elucidation of the mechanism of its replication²⁶.

We thank Elizabeth Hennessey, Eureka Espanol, Karen Bowman, Darlene Groth, Patricia Cochran and Michael Braunfeld for technical assistance; Rebecca Mead, Susan Mailot, Lorraine Gallagher and Frances Elvin for editorial and administrative assistance and Drs George Lewis, Michael McKinley, David Bolton and Michael Oldstone for discussion. P.E.B. is a postdoctoral fellow of the National Multiple Sclerosis Society. Research and training support from the N.I.H. (AG 02132, NS 14069 and NS 07219) and funds from R. J. Reynolds Industries, Inc., Sherman Fairchild Foundation and W. M. Keck Foundation are acknowledged.

Received 2 April; accepted 1 June 1984.

1. Zlotnik, I. & Rennie, J. C. *J. comp. Path.* **75**, 147–157 (1965).
2. Beck, E. & Daniel, P. M. in *Slow Transmissible Diseases of the Nervous System* Vol. 1 (eds Prusiner, S. B. & Hadlow, W. J.) 253–270 (Academic, New York, 1979).
3. McFarlin, D. E., Raff, M. C., Simpson, E. & Nehlsen, S. *Nature* **233**, 336 (1971).
4. Gibbs, C. J. Jr *Curr. Top. Microbiol. Immun.* **40**, 44–58 (1967).
5. Kingsbury, D. T., Smeltzer, D. A., Gibbs, C. J. Jr & Gajdusek, D. C. *Infect. Immun.* **32**, 1176–1180 (1981).
6. Stites, D. P., Garfin, D. E. & Prusiner, S. B. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 211–224 (Academic, New York, 1979).

7. Kasper, K. C., Bowman, K., Stites, D. P. & Prusiner, S. B. in *Hamster Immune Responses in Infectious and Oncologic Diseases* (eds Streilein, J. W., Hart, D. A., Stein-Streilein, J., Duncan, W. R. & Billingham, R. E.) 401–413 (Plenum, New York, 1981).
8. Kasper, K. C., Stites, D. P., Bowman, K. A., Panitch, H. & Prusiner, S. B. *J. Neuroimmun.* **3**, 187–201 (1982).
9. Prusiner, S. B. *et al. Biochemistry* **21**, 6942–6950 (1982).
10. Bolton, D. C., McKinley, M. P. & Prusiner, S. B. *Science* **218**, 1309–1311 (1982).
11. McKinley, M. P., Bolton, D. C. & Prusiner, S. B. *Cell* **35**, 57–62 (1983).
12. Prusiner, S. B. *et al. Cell* **35**, 349–358 (1983).
13. Diringer, H. *et al. Nature* **306**, 476–478 (1983).
14. Hawkes, R., Niday, E. & Gordon, J. *Analyt. Biochem.* **119**, 142–147 (1982).
15. Burnette, W. N. *Analyt. Biochem.* **112**, 195–203 (1981).
16. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
17. Sternberger, L. A. in *Immunocytochemistry* 2nd edn, 104–169 (Wiley, New York, 1979).
18. Beck, E. & Daniel, P. M. in *Slow, Latent and Temperate Virus Infections*, NINDB Monogr. 2 (eds Gajdusek, D. C., Gibbs, C. J. Jr & Alpers, M.) 203–206 (U.S. Government Printing Office, Washington DC, 1965).
19. Bruce, M. E., Dickinson, A. G. & Fraser, H. *Neuropath. appl. Neurobiol.* **2**, 471–478 (1976).
20. Baringer, J. R., Bowman, K. A. & Prusiner, S. B. *J. Neuropath. exp. Neurol.* **40**, 329 (1981).
21. Tateishi, J., Nagara, H., Hikita, K. & Sato, Y. *Ann. Neurol.* **15**, 278–280 (1984).
22. Mountjoy, C. Q., Tomlinson, B. E. & Gibson, P. H. *J. neurol. Sci.* **57**, 89–103 (1982).
23. Gajdusek, D. C. *Science* **197**, 943–960 (1977).
24. Hadlow, W. J., Prusiner, S. B., Kennedy, R. C. & Race, R. E. *Ann. Neurol.* **8**, 628–631 (1980).
25. Prusiner, S. B. & Kingsbury, D. T. *CRC Crit. Rev. clin. Neurobiol.* (in the press).
26. Prusiner, S. B. *Science* **216**, 136–144 (1982).
27. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
28. Sigel, M. B., Shina, Y. N. & Vanderlaan, W. P. in *Meth. Enzym.* **93**, 3–12 (1983).
29. Bolton, A. E. & Hunter, W. M. *Biochem. J.* **133**, 529–539 (1973).
30. Marsh, R. & Kimberlin, R. *J. infect. Dis.* **131**, 104–110 (1975).
31. Puchtler, H., Sweaf, F. & Levin, M. *J. Histochem. Cytochem.* **10**, 355–364 (1962).

Localization of a T-cell receptor diversity-region element

Joshua Kavalier, Mark M. Davis & Yueh-hsiu Chien

Department of Medical Microbiology, Stanford University Medical School, Stanford, California 94305, USA

Recently, complementary DNA clones encoding one chain of the T-cell receptor for antigen have been isolated from both murine and human cell lines^{1–3}. Sequence analysis of these clones indicates that they encode elements analogous to the variable (V), constant (C) and joining (J) regions of immunoglobulins^{2,3} and that this corresponds to the β -chain subunit of the T-cell receptor complex⁴. These genes are rearranged in the genomes of specific T-cell lines and hybrids but not in other cell types¹. Analysis of the components of one such rearranged gene, 2B4, isolated from the pigeon cytochrome c-specific, H-2E-restricted T helper (T_H) hybridoma⁵, and its unrearranged (liver) counterpart, indicate that an 8-nucleotide sequence 3' to the rearranged variable region is not derived from either the germ-line V- or J-region gene segments⁶. As this sequence lies at a similar position to the diversity (D) region in immunoglobulin heavy-chain genes⁷, we postulated the existence of an array of germ-line D-region elements that would contribute significantly to the number of different β -chain molecules which could be created. Here we describe the localization and sequence of one such D-region element, found ~650 nucleotides 5' to the first J_T cluster⁸. This element seems to be involved in both functional (V–D–J) and non-functional (D–J) rearrangements. Our observations, combined with previous results^{6,8,9}, indicate that variable-region formation (V–D–J joining) in the T-cell receptor β -chain gene follows the 12/23-base pair (bp) rule of rearrangement established for the recombination of immunoglobulin gene segments^{7,10}, but that the organization of the heptamer and non-amer element found surrounding the D region is significantly different.

The cDNA clone used in these studies, 86T5, was one of a series of three clones isolated from a BALB/c thymocyte library². It contains the same constant region (C_T) and joining region (J_T)⁸ as found in the apparently functional 86T1 cDNA clone², but the 80 nucleotides 5' to the J element do not encode any of the highly conserved amino acid sequences found encoded in most V_T genes at this position^{2,3,6}. Characterization of the constant-region locus of the murine β -chain, using 86T5 as a probe on Southern blots of the relevant genomic clones, revealed hybridization to two distinct regions apart from the constant-region exons: a fragment subsequently shown to contain the

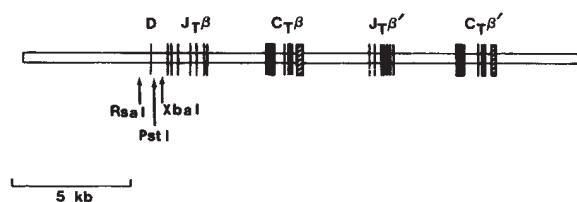


Fig. 1 The sequence organization of *D*, *J* and *C* regions. B10.A liver genomic clones were isolated and characterized with respect to the *J* and *C* elements as described previously^{6,8}. Restriction sites relevant to Fig. 2a are indicated.

unrearranged *J*_T3 element as expected, and an unexpected strong hybridization some distance 5' to *J*_T1.

Figure 1 shows the previously characterized *C*- and *J*-region loci^{6,8}, together with the *D*-region element now described. The location of this *D* region is very similar to that of an immunoglobulin *D* (Q52) which lies 700 nucleotides 5' to the heavy-chain *J* cluster¹¹. Details of the sequencing strategy for this 5' region are shown in Fig. 2a. The sequence (Fig. 2b) contains a *D*-region element flanked on either side by conserved 7 and 9 nucleotide sequences homologous to those of immunoglobulin (see Table 1) which are thought to be the signal sequences for recombination¹⁰. As indicated in Fig. 2b, this portion of the DNA contains all of the 5' 80 nucleotides of 86T5, diverging from the cDNA clone just before *J*_T3. Directly 3' to the point of divergence in the germ-line DNA is the sequence CACGGTG-[23 nucleotides]-ACAAAAACC, which differs by one nucleotide from the consensus sequence 3' to the *D* regions of immunoglobulin (Table 1)¹⁰ and is identical to that of Q52 (ref. 11). Immediately 5' to the *D* region is the sequence CTTTTTGT-[12 nucleotides]-CATTGTG, similar to the signal sequences found 5' both to the *J*_T elements and to *D*_H regions in the immunoglobulin heavy chain genes. In 86T5, the sequence of the nonamer differs from this by one nucleotide (Fig. 2b), perhaps due to a strain polymorphism (86T5 is from BALB/c and the genomic clones are from B10.A). As we could find no V-like sequence on the 5' side of the recombination site (the highly conserved residues 90-94 are missing^{2,3,6}) and as there is no consensus RNA splicing sequence¹² on the 3' side of the divergence, we conclude that 86T5 is the transcript of a chromosomal rearrangement that involved the joining of *D* and *J* elements.

This seems to be a functional *D* region, as it is also used in the 86T1 (ref. 2) and 86C6C (unpublished data) cDNA clones, which have complete *V-D-J-C* coding blocks. The homologies to the 86T1 and 86C6 sequences are shown in Table 1. Therefore

Table 1 Nucleotide sequence comparison of T-cell receptor and immunoglobulin

3' Flanking	Heptamer	Spacer (-bp)	Nonamer
<i>V_k</i>	CACAGTG	12 ± 1	GGTTTTGT
<i>V_λ</i>	CACAATG	23	GGTTTTGC
<i>V_H</i>	CACAGTG	23 ± 1	NGTTTTGT
<i>V_{Tβ}</i>	CACAGCA	23	AGTTTTGT
<i>D_H</i> (3' side)	CACAGTG	12	GATTTTGT
<i>D_T</i> (3' side)	CACGGTG	23	CTTTTTGT
5' Flanking	Nonamer	Spacer (bp)	Heptamer
<i>J_k</i>	ACAAAAACC	23 ± 1	CACTGTG
<i>J_{λ1}</i>	ACAGAACA	12	CACAGTG
<i>J_H</i>	ACAAAAACC	23 ± 1	NACTGTG
<i>J_{Tβ}</i>	GCATAAACC	12 ± 2	NGCTGTG
<i>D_H</i> (5' side)	ACAAAAACC	12	TACTGTG
<i>D_T</i> (5' side)	ACAAAAACC	12	CATTGTG

Germ-line <i>D</i>	
86T1	GGACAGGGGG
86T5	GGACAGGGGG
86C6	GGACAGGA
	GGACAG

this element is involved in both functional and non-functional rearrangements. As with immunoglobulins¹⁰, the point of recombination between *D* and *J* is variable.

The spacings of the signal sequences for DNA rearrangement in the T-cell receptor β-chain *V* region (*V-D-J*) closely follows the 12/23-bp rule first postulated for immunoglobulins by Early *et al.*⁷. Table 1 indicates the approximate consensus sequences for the immunoglobulin and T-cell receptor heptamer and nonamer as well as the spacing in between, and reveals the remarkable similarities between these sequences; in particular, the 3'-flanking regions of *D_H* and *D_T* are almost identical. However, the sequences 3' to the *V* regions and 5' to *D* are more dissimilar (2/7 and 2/9 mismatches) and might result in some regulatory differences, as previously noted for the *J_T* flanking sequences^{6,8}.

This similarity of the heptamer and nonamer sequences as well as the preservation of the 12/23-bp rule indicates that identical or nearly identical mechanisms must be operating in the rearrangement of these gene segments and those of immunoglobulin. However, as shown in Fig. 3, there are fundamental differences in the arrangement of these signal sequences between immunoglobulin and the T-cell receptor β-chain. In particular, if the joining process requires no other recognition

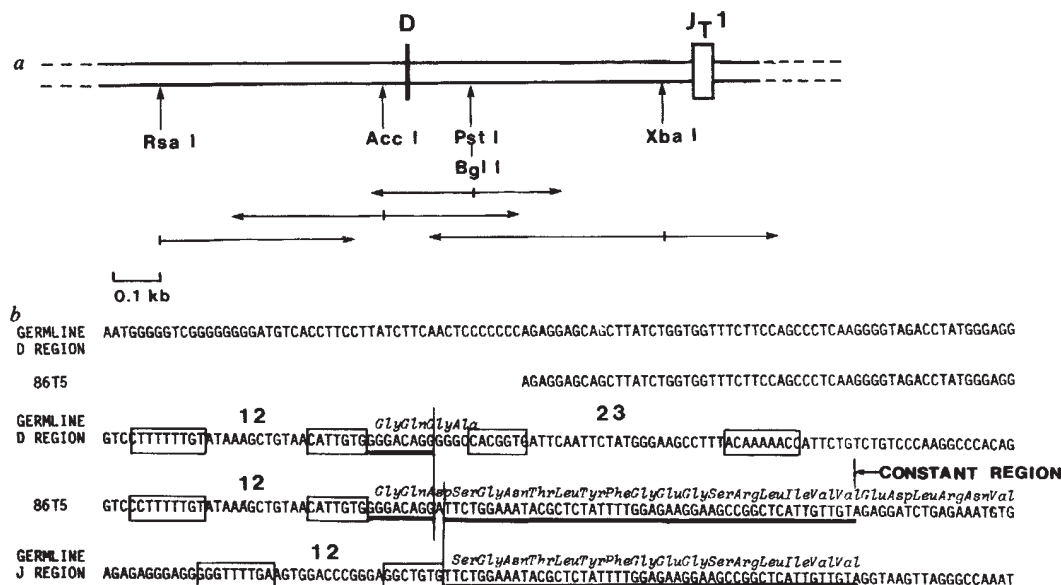


Fig. 2 a, Sequence strategy in the *D* region. A subclone encompassing this region has been described previously⁸. Vertical arrows indicate the sites that were labelled with polynucleotide kinase¹⁸. Sequencing was performed by the procedures of Maxam and Gilbert¹⁸. **b**, Comparison of a germ-line *D* (present study) and *J*_T3 region⁸ sequences with the 86T5 sequence². Coding regions have the translated amino acid sequence shown above. Homologies are denoted by solid lines. Boxes enclose the conserved heptamer and nonamer signal sequences and the distances between them are indicated.

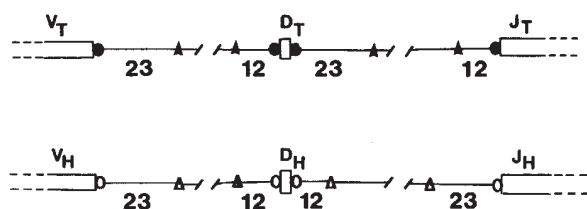


Fig. 3 Arrangement of the signal sequences flanking germ-line T-cell receptor segments and germ-line immunoglobulin heavy-chain segments. Circles represent the conserved heptamers; triangles represent the conserved nonamers.

sequences, it is formally possible to have V_T joining directly to J_T , as well as a given D_T element joining to another D_T element. The first possibility would result in a smaller V region, the second in a larger one. This might substantially increase the diversity of T-cell receptor binding sites. Overall, however, the observed similarities between this chain of the T-cell receptor and immunoglobulins (reviewed in ref. 13) clearly show the common origins of the respective gene families and indicates that this rearrangement mechanism probably occurred only once in the course of evolution. Another indication of the shared machinery for recombination between B cells and T cells are the relatively frequent (10–50%) instances of aberrant rearrangement of the immunoglobulin heavy-chain locus in T cells. In particular, Kurosawa *et al.* characterized several D – J joining events in T-cell tumours¹⁴. This phenomenon might now be explained by postulating that enzymes activated to rearrange T-cell receptor genes are able to act, at least partially, on immunoglobulin genes.

The adenosine residue adjacent to the recombination site of D and J of 86T5 (underlined in Table 1) was not accounted for by either the germ-line J or D sequences. This does not seem to be due to strain differences because adenosine is not present in the 86T1 D -region sequence. This is similar to the finding that, in immunoglobulin heavy-chain gene rearrangement, extra nucleotides (N regions) are inserted between D and J_H as well as between V and D as part of the joining process¹⁵. Most of the D regions we have studied are rich in dG residues at the boundaries of V – D and D – J joints on the coding strand (unpublished observation). Such G tracts are often seen with the immunoglobulin D regions but are not found in their germ-line counterparts¹⁵. However, in the D region described here, all the G residues can be accounted for in the germ-line sequence and thus it is not necessary to evoke an enzymatic mechanism¹⁵.

The finding that 86T5 is derived from a non-productive rearrangement and the earlier result that a second thymocyte-derived cDNA clone, 86T3, is a partial readthrough product of the constant region⁸ indicates that two out of the three thymocyte-derived clones (86T1, 86T3 and 86T5, ref. 2) are non-functional. This ratio could be higher, as 7 of the original 10 cDNA clones from the thymocyte library were not characterized further because they had small inserts, similar in size to 86T5 (S. Hedrick & M.M.D., unpublished). This apparent failure rate of the β -chain gene in thymocytes is interesting in view of the finding that a relatively small fraction (perhaps 15%) of human thymocytes have detectable receptor molecules on their surface¹⁶, despite the fact that most cells in the thymus express this gene in their RNA (P. Fink and S. Hedrick, manuscript in preparation). One explanation of this is that two independent types of selection are occurring in the thymuses of immature animals; the first being the production of a viably rearranged gene at both the α and β loci and the second the selection for self-compatibility with the major histocompatibility complex. The combination of these two selections may account for the considerable cell death (>99%) that seems to occur in this organ¹⁷.

We thank N. R. J. Gascoigne for helpful comments on the manuscript and K. Redman for assistance in its preparation. This work was funded by grants to M.M.D. from the NIH (AI19512 and AI20320) and from the American Cancer Society (IM00364).

Note added in proof: Further analysis of the β -chain locus discussed here indicates the existence of a second D region element 540 nucleotides upstream of the 3' cluster of J regions (D. M. Becker *et al.*, work in progress). This D region exhibits the same 12/23 spacing of heptamer and nonamer elements as the one described, and appears in the TM86 cDNA clone. Similar data have been obtained by M. Kronenberg and co-workers (personal communication).

Received 29 May; accepted 27 June 1984.

- Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. *Nature* **308**, 149–153 (1984).
- Hedrick, S. M., Nielson, E. A., Kavalier, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 153–158 (1984).
- Yanagi, Y. *et al.* *Nature* **308**, 145–149 (1984).
- Acuto, O. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 3851–3855 (1984).
- Samelson, L. E., Germain, R. N. & Schwartz, R. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6972–6976 (1983).
- Chien, Y., Gascoigne, N. R. J., Kavalier, J., Lee, N. E. & Davis, M. M. *Nature* **309**, 322–326 (1984).
- Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. *Cell* **19**, 981–992 (1980).
- Gascoigne, N. R. J., Chien, Y., Becker, D. M., Kavalier, J. & Davis, M. M. *Nature* **310**, 387–391 (1984).
- Siu, G. *et al.* *Cell* **37**, 393–401 (1984).
- Tonegawa, S. *Nature* **302**, 575–581 (1983).
- Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. *Nature* **290**, 562–565 (1981).
- Mount, S. M. *Nucleic Acids Res.* **10**, 459–472 (1982).
- Davis, M. M., Chien, Y., Gascoigne, N. R. J. & Hedrick, S. M. *Immun. Rev.* (in the press).
- Kurosawa, Y. *et al.* *Nature* **290**, 565–570 (1981).
- Alt, F. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118–4122 (1982).
- Acuto, O. *et al.* *Cell* **34**, 717–726 (1983).
- Scollay, R., Butcher, E. & Weissman, I. *Eur. J. Immun.* **10**, 210–217 (1980).
- Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).

Nucleotide sequence comparison of normal and translocated murine *c-myc* genes

Lawrence W. Stanton, Paul D. Fahrlander,
Paul M. Tesser & Kenneth B. Marcu

Department of Biochemistry and Molecular Biology Graduate Program, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

Plasmacytoid tumours in mice and Burkitt's lymphomas in humans display characteristic chromosomal translocations involving the *c-myc* proto-oncogene^{1,2}. Models to explain quantitative changes in *c-myc* expression have been proposed based on the loss of normal promoters^{3–9} as a result of translocation. However, alternative explanations^{10–16}, such as somatic mutation^{14,17,18}, are needed to explain altered *c-myc* expression in the absence of gene breakage. We present here the nucleotide sequence of the normal murine *c-myc* gene. Comparison of this sequence with that of a translocated *c-myc* gene from a murine plasmacytoma reveals complete identity of coding sequence. One nucleotide difference was found in the non-coding first exon. This shows that qualitative changes of the *c-myc* gene product are not required for oncogenesis in murine plasmacytomas. In contrast, mutations are found in coding¹⁸ and non-coding¹⁴ regions of translocated *c-myc* genes from Burkitt's lymphomas, suggesting that the mechanisms by which *c-myc* is activated in plasmacytomas and Burkitt's lymphomas are different.

Although it is well established that the *c-myc* gene is rearranged as a result of chromosomal translocations in most plasmacytomas, it remains uncertain whether rearrangements of this sort are sufficient for *c-myc* to confer its oncogenic potential. To determine whether the translocated, rearranged *c-myc* genes necessarily contain mutations, the nucleotide sequence of a normal murine *c-myc* cDNA was compared with the corresponding sequence of a rearranged *c-myc* oncogene cloned from a murine plasmacytoma. The normal *c-myc* cDNA

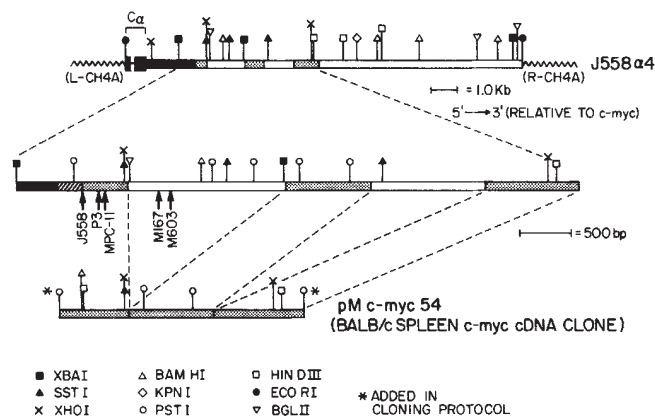


Fig. 1 Restriction maps of molecularly cloned *c-myc* genes. Clone J558 α 4 was isolated from a genomic library of the J558 solid tumour¹⁹ and represents the translocated *c-myc* gene recombined near the immunoglobulin C α gene (solid blocks). pM-c-myc54 is a *c-myc* cDNA clone isolated from a BALB/c spleen cDNA library⁶. Positions of intron (open blocks) and exon (stippled blocks) regions are indicated. Arrows indicate the breakpoints within the *c-myc* gene from other murine plasmacytomas^{6,24,27-30}.

clone, pM-c-myc 54, was isolated from a BALB/c spleen cDNA library⁶; the translocated *c-myc* gene is from the BALB/c plasmacytoma, J558¹⁹, and is transcribed at elevated levels in these cells³. Figure 1 shows maps of the clones. Appropriate regions of the clones were sequenced by the chain termination technique²⁰ using cloned restriction fragments in M13 vectors mp8, mp9, mp10 and mp11²¹. Certain cloned restriction fragments were too large to allow sequencing of the entire insert and in such instances, M13 deletions were generated using *Bal*31²² or DNase I²³ methods of directed deletion cloning. This allowed accurate sequence determination of the *c-myc* gene in regions which are devoid of convenient cloning sites.

The complete nucleotide sequence of the pM-c-myc 54 cDNA clone is presented in Fig. 2. The clone lacks 19 base pairs (bp) at its 5' end, assuming that the 5'-most promoter and capping sequences are used^{9,24}. A loss of some nucleotides from the 5' end of a cDNA clone is a common artefact of this cDNA cloning procedure²⁵. The clone contains a 376-bp 3' non-coding sequence but lacks some 3'-terminal nucleotides, as no poly(A) tail is present. A poly(A) recognition sequence, AAUAAA, 51 bp from the 3' end of the clone was not used by the *c-myc* mRNA. Because the cDNA clone is 2,252 bp long, and the largest observed size for the normal *c-myc* mRNA is 2.4 kilobases (kb), we conclude that little 3' non-coding sequence is absent, assuming an average poly(A) tail length of about 150 nucleotides. The first exon contains multiple stop codons in every reading frame and therefore represents non-translated sequences. The first ATG codon is at the 5' end of the second exon (see Fig. 2). This ATG is followed by a long, open reading frame, suggesting that it is the initiation point for translation. Figure 2 also gives the predicted amino acid sequence of the *c-myc* protein. The translocation break points from several plasmacytomas are indicated and do not affect the amino acid coding portion of the *c-myc* gene.

Determination of the exon sequences of a translocated *c-myc* gene cloned from J558 α 4, a BALB/c J558 plasmacytoma¹⁹ and comparison with the normal *c-myc* cDNA sequence reveals that the two clones show complete sequence identity except at position 235. This residue, which is in the non-coding first exon, is A in the cDNA and G in J558 α 4. Interestingly, this nucleotide is also G in another translocated *c-myc* gene that was isolated from the BALB/c plasmacytoma MPC-11²⁴, and a previously reported nucleotide sequence of the translocated *c-myc* gene from J558 also has a G residue at this position⁹. The significance of this single difference is unknown. It may represent a polymorphism in the BALB/c *c-myc* locus. Alternatively, the change may have occurred within the cDNA clone as a result of the cloning procedure either *in vitro* or during clonal propagation in bacteria.

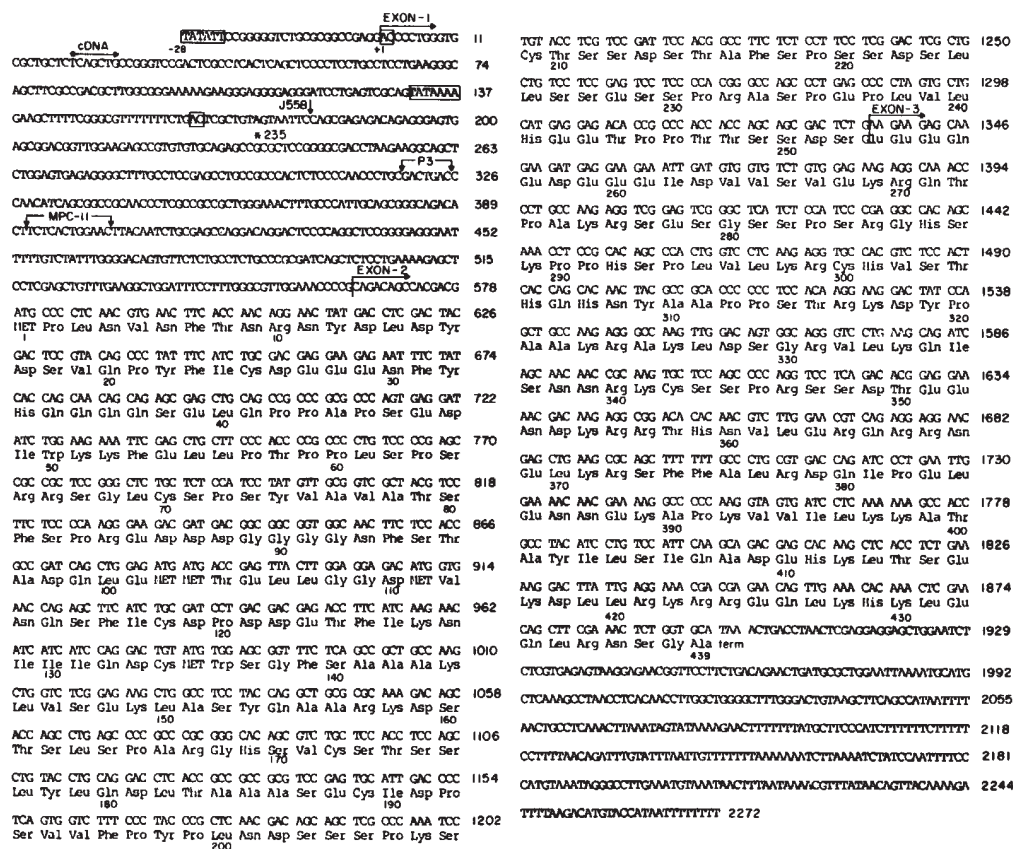


Fig. 2 Nucleotide sequence of murine *c-myc* cDNA, pM-c-myc 54. An additional 47 nucleotides located 5' to the beginning of the cDNA²⁴ are included to present the entire transcription unit. Two TATA boxes and cap sites are boxed. The predicted amino acid sequence is shown above the proposed coding region. Vertical arrows indicate precise break points for other *c-myc* translocation^{6,24,28-30}. The start of each exon is shown by over-head arrows. Position 235 is the only nucleotide different from the sequence of the translocated *c-myc* gene, J558 α 4. There are a few differences between this sequence and another published J558 rearranged *c-myc* gene sequence⁹. The reasons for these minor discrepancies remain unclear but they do not result in qualitative changes in coding potential of the rearranged *c-myc* gene. Such differences may reflect the origin of these rearranged *c-myc* genes. Our J558 *c-myc* clone was from the subcutaneous tumour¹⁹ whereas the latter gene was isolated from the J558L tissue culture cell line²⁸.

Perhaps more significant in the context of the present discussion is the complete identity of the amino acid coding portions of *c-myc* genes in their normal and translocated states. The lack of differences supports the hypothesis that the *c-myc* gene product is itself qualitatively normal in plasmacytomas. It is likely that an inappropriate, quantitative change in the normal gene product is the common lesion in these malignancies. Quantitative changes in *c-myc* expression could follow from disruption of control and promoter sequences as a result of translocation or translocation to regions of greater transcriptional activity such as open chromatin or enhancer element-activated regions. Although both of these explanations involve transcriptional control of *c-myc* expression, regulation at the level of translation may also be important^{6,26}. Additional sequence data concerning the normal *c-myc* gene may clarify the significance of the solitary nucleotide difference described here. In any event, the extensive mutations observed in the translocated *c-myc* loci of Burkitt's lymphomas^{14,18} are clearly not observed in the analogous translocation of the plasmacytomas. This conclusion is further supported by S₁ nuclease analyses performed on the *c-myc* RNAs expressed in other translocation-positive plasmacytomas¹⁶. The *c-myc* translocation in Burkitt's lymphoma may occur earlier in B-cell ontogeny than its murine counterpart. This could make the translocated *c-myc* allele in Burkitt's lymphoma more susceptible to the somatic mutator acting on a nearby immunoglobulin gene.

We thank Ms Mary Vogelle for assistance and Ms Mary Ann Huntington for help in preparation of this manuscript. This research was supported by NIH grants GM26939 and AI00416 and American Cancer Society (ACS) grants NP-405 and RD-

178. K.B.M. is a Research Career Development Awardee of the NIH. P.D.F. is a postdoctoral fellow of the ACS.

Received 9 April; accepted 8 June 1984.

1. Klein, G. *Cell* **32**, 311-315 (1983).
2. Perry, R. P. *Cell* **33**, 647-649 (1983).
3. Marcu, K. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 519-523 (1983).
4. Mushinski, J. F., Bauer, S. R., Potter, M. & Premkumar-Reddy, E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1073-1077 (1983).
5. Erikson, J., Ar-Rushdi, A., Orwina, H. L., Vowell, P. C. & Croce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 820-824 (1983).
6. Stanton, L. W., Watt, R. & Marcu, K. B. *Nature* **303**, 401-406 (1983).
7. Hayday, A. C. *et al. Nature* **307**, 334-340 (1984).
8. Adams, J., Gerondakis, S., Webb, E., Corcoran, L. M. & Cory, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1982-1986 (1983).
9. Bernard, O., Cory, S., Gerondakis, S., Webb, E. & Adams, J. M. *EMBO J* **2**, 2375-2383 (1983).
10. Leder, P. *et al. Science* **222**, 765-771 (1983).
11. Nishikura, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4822-4826 (1983).
12. Croce, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6922-6926 (1983).
13. Erikson, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7581-7585 (1983).
14. Taub, R. *et al. Cell* **36**, 339-348 (1984).
15. Hollis, G. F. *et al. Nature* **307**, 752-755 (1984).
16. Yang, J. Q. *et al. in Current Topics in Microbiology and Immunology* (eds Potter, M., Melchers, F. & Weigert, M.) (Springer, New York, in the press).
17. Hamlyn, P. H. & Rabbitts, T. H. *Nature* **304**, 135-139 (1983).
18. Rabbitts, T. H., Hamlyn, P. H. & Baer, R. *Nature* **306**, 760-765 (1983).
19. Harris, L. J., Land, R. B. & Marcu, K. B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4175-4179 (1982).
20. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
21. Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309-322 (1981).
22. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4298-4302 (1982).
23. Frischhauf, A., Garoff, H. & Lebrach, H. *Nucleic Acids Res.* **8**, 5541-5549 (1980).
24. Stanton, L. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 829-835 (1984).
25. Watt, R. *et al. Nature* **303**, 725-728 (1983).
26. Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7476-7480 (1983).
27. Calame, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6994-6998 (1982).
28. Adams, J. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6966-6970 (1982).
29. Neuberger, M. S. & Calabi, F. *Nature* **305**, 240-243 (1983).
30. Dunnick, W., Shell, B. E. & Dery, C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7269-7273 (1983).

Molecular cloning of a human DNA repair gene

A. Westerveld*, J. H. J. Hoeijmakers, M. van Duin, J. de Wit, H. Odiijk, A. Pastink†, R. D. Wood‡§ & D. Bootsma

Department of Cell Biology and Genetics, Erasmus University, Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands
† Department of Medical Biochemistry, State University of Leiden, 2333 AL Leiden, The Netherlands
‡ Lawrence Berkeley Laboratory, Biomedical Division and Graduate Group in Biophysics, University of California, Berkeley, California 94720, USA

Cell strains derived from patients having a hereditary disorder associated with defects in repair of DNA damage such as xeroderma pigmentosum^{1,2} and mutants isolated from established rodent cell lines^{3,4} provide the tools for genetic and biochemical analysis of DNA repair pathways in mammalian cells. Complementation studies using these cells have illustrated the genetic and biochemical complexity of these pathways^{3,5,6}. The precise nature of the genes and gene products involved in these mutants has not yet been resolved. Isolation of repair genes by recombinant DNA technology would open up new approaches to the elucidation of repair mechanisms in mammalian cells. Here we report the molecular cloning of a human repair gene (*ERCC1*) that complements the repair defect in a Chinese hamster ovary (CHO) mutant cell line.

The CHO mutant cell line 43-3B (ref. 7) used in this study is sensitive to UV light and mitomycin C and falls into complementation group 2 of the classification described by

Thompson *et al.*^{3,7}. The mitomycin C sensitivity of these cells allows efficient selection of repair-competent transformants after transfection with human DNA⁸.

The strategy for the isolation of the human repair gene is depicted in Fig. 1. High-molecular weight DNA isolated from HeLa cells is partially cleaved to an average fragment size of 50-60 kilobase pairs (kbp), ligated to linearized plasmid pSV3gptH and transfected to the 43-3B CHO mutant cells. pSV3gptH contains the *Escherichia coli gpt* (*Ecogpt*) gene that renders transformed cells resistant to medium containing mycophenolic acid (MPA medium⁹), which provides a means of preselecting the fraction of transfection-competent recipient cells¹⁰. The physical linkage between pSV3gptH and the repair gene can be used to identify in an *E. coli* recombinant library of a (secondary) transformant, clones that contain the two genes, using colony hybridization with a labelled *Ecogpt* probe¹¹.

The transfection frequency of 43-3B cells with the human-pSV3gptH hybrid DNA molecules for MPA resistance (MPA^r) was 50 clones per µg DNA per 10⁶ cells and for combined MPA^r + UV^r or MPA^r + mitomycin C resistance (MM-C^r) 0.02 clones per µg DNA per 10⁶ cells. The latter transformation efficiency is comparable to that found for another CHO mutant of the same complementation group⁸ and is in the order of that expected for unique genes^{12,13}. One UV^r and one MM-C^r transformant (designated 43IUV and 43IMM-C respectively) were analysed for the presence of pSV3gptH and human sequences by blot hybridization and characterized with respect to various repair parameters. Southern blot analysis revealed the presence in both genomes of multiple copies (probably ≥100) of the dominant marker as well as a considerable amount of human DNA (see, for 43IUV, Fig. 2a, e). As a first end point of repair, we investigated UV-induced unscheduled DNA synthesis in the transformants. The level of unscheduled DNA synthesis in mutant 43-3B is only a fraction of that performed by repair-proficient CHO cells, while the level in transformants 43IUV and 43IMM-C equals that of the CHO wild type (Table 1). As in other rodent cells, the level of unscheduled DNA synthesis in CHO wild-type cells is lower than that in human cells¹⁴.

* To whom reprint requests should be addressed.

§ Present address: Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, PO Box 6666, New Haven, Connecticut 06511, USA.

Survival curves show that both transformants have regained the repair-proficient constitution for UV and MM-C (Fig. 3). The finding that these curves do not coincide with those of CHO wild-type cells might be explained by the heterologous (human) nature of the gene and its product. As repair-competent revertants were never observed in parallel mock-transfection experiments, we conclude that the repair-proficient character of 43IUUV and 43IIMM-C is due to the stable incorporation and expression of a complementing human repair gene.

After transformation of 43-3B cells with 43IUUV DNA (secondary transfection), coupled transfer of MPA^r and MM-C^r was found to be only slightly lower than that of either marker alone (about 0.02 clones per μg DNA per 10^6 cells), indicating a close physical linkage between the two in the DNA of 43IUUV. One secondary transformant, 43IIMM-C, was selected for further characterization. The levels of its unscheduled DNA synthesis and its MM-C- and UV-survival curves were not significantly different from those of the primary transformant (Fig. 3, Table 1). Southern blot analysis of the DNA of 43IIMM-C revealed that three copies of the dominant marker and a specific subset of human DNA fragments had been retained (Fig. 2a, e). Restriction enzyme mapping data indicated that two of the three pSV₃gptH copies were very closely linked, whereas the third copy was located at some distance from the other two (see below). In a tertiary transformation, combined transfer of MPA^r and MM-C^r was found with a similar frequency as for the secondary transformation, indicating that no rearrangement had

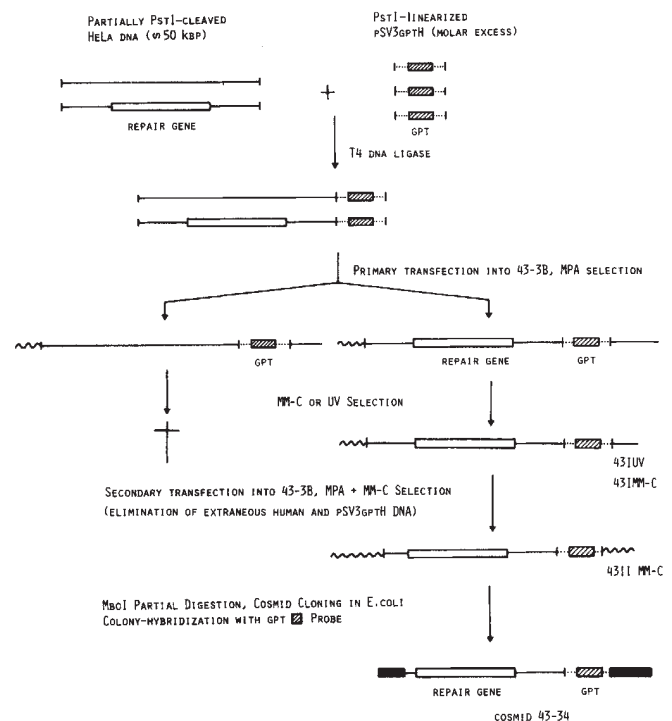
disconnected the two genes. Therefore DNA of the secondary transformant was used to construct a cosmid recombinant library in *E. coli* (Fig. 1).

Restriction endonuclease analysis of cosmids of seven *E. coli* clones hybridizing with the *Ecogpt* probe revealed that all had overlapping sequences (for a representative trio covering the entire 70-kbp cloned region, see Fig. 2b-d). Some harboured all three pSV₃gptH copies present in the 43IIMM-C genome (cosmid 43-31), others contained only two (cosmid 43-12) or one copy (cosmid 43-34, Fig. 2b). The remaining part of the inserts consisted of only human DNA. From the hybridization with human repeats we conclude that most of the human sequences integrated in 43IIMM-C are covered by these clones (compare Fig. 2d and e). A physical map of the cosmid inserts (Fig. 4, upper part) appeared to be consistent with the Southern blot data obtained from the genomic DNA of 43IIMM-C. Apparently this transformant has stably integrated a continuous 43IUUV DNA fragment longer than 70 kbp.

After transfection to 43-3B all cosmids induced MPA resistance with a high frequency (70 clones per μg DNA per 10^6 cells); however, only cosmid 43-34 transferred resistance to MM-C and UV (Fig. 3) and restoration of unscheduled DNA synthesis (Table 1) with an efficiency comparable to that of the dominant marker. This is more than 3×10^3 higher than that of genomic DNA. Cosmid 43-34 extends the furthest to the right in the map of Fig. 4, and contains only one pSV₃gptH copy, indicating that the UV/MM-C-resistance gene must reside to

Fig. 1 Experimental strategy for the isolation of the human repair gene *ERCC1*.

Methods: For transfections, $\sim 5 \times 10^5$ 43-3B cells were seeded in 100 mm Petri dishes 1 day before DNA transformation. The cells were grown in F10/Dulbecco's minimal essential medium (DMEM) 1:1 supplemented with antibiotics and 3% fetal and 7% newborn calf serum. High-molecular weight HeLa DNA, isolated as described previously¹⁷, was partially cleaved to an average fragment size of 50–60 kbp with restriction endonuclease *Pst*I. Provided that the repair gene is not exceptionally large (>40 –50 kbp), sufficient gene copies in the DNA should remain intact. The restricted DNA was subsequently ligated to a twofold molar excess of *Pst*I-linearized dominant marker pSV₃gptH using T4 DNA ligase (Biolabs). Vector pSV₃gptH is a derivative of pSV₃gpt (ref. 9), from which the simian virus 40 (SV40) early region and a 120-bp *Hind*III–*Bgl*II fragment in front of the *gpt* sequence (which contains false start codons¹⁸) were removed. After testing the ligation on agarose gels the hybrid molecules were transfected into 43-3B CHO repair mutant cells by the calcium phosphate precipitation method¹⁹. 20 μg of DNA (vector + HeLa DNA) were applied to each Petri dish. Following overnight exposition of the DNA, cells were treated with dimethyl sulphoxide (10% for 30 min) and grown for 24–48 h on non-selecting culture medium to allow expression of the transfected markers. Selection for the dominant marker was in modified MPA medium containing: F10/DMEM 1:1, antibiotics (3% fetal, 7% newborn calf serum), aminopterin (0.2 $\mu\text{g ml}^{-1}$), thymidine (5 $\mu\text{g ml}^{-1}$), xanthine (10 $\mu\text{g ml}^{-1}$), hypoxanthine (15 $\mu\text{g ml}^{-1}$), mycophenolic acid (25 $\mu\text{g ml}^{-1}$) and deoxycytidine (2.3 $\mu\text{g ml}^{-1}$). The selection medium was refreshed every 3–4 days. After the appearance of MPA-resistant (MPA^r) colonies (within 2 weeks) the cells of each Petri dish were reseeded on two dishes, one of which was UV-irradiated ($3 \times 5 \text{ J m}^{-2}$ at 1-day intervals), the other exposed to mitomycin C (10^{-8} M)⁸ in MPA medium. These treatments were lethal to the 43-3B mutant, but not to repair-competent CHO cells. The UV^r and MM-C^r colonies appearing on the duplicates of one parental Petri dish were expanded into mass culture (10^8 – 10^9 cells) in selective conditions (except for UV) and characterized with respect to UV and mitomycin C sensitivity (Fig. 3), UV-induced unscheduled DNA synthesis (Table 1) and Southern blot hybridization (Fig. 2). Secondary and tertiary transformation and selection were done as above except that 43IUUV and 43IIMM-C DNA, respectively, were used as donor DNAs and that selection for both markers (MPA^r and MM-C^r) was done simultaneously. To construct a cosmid library of 43IIMM-C DNA, a partial *Mbo*I size-fractionated digest of 43IIMM-C DNA with an average fragment size of 40–50 kbp was ligated to *Bam*HI-cleaved pTCF-cosmid vector arms, packaged *in vitro* and transduced into bacterial host ED8767 essentially as described earlier²⁰. This library, which consisted of 0.8×10^6 independent recombinants (equivalent to five times the diploid Chinese hamster genome), was screened by colony filter hybridization^{20,21} using an *Ecogpt* probe (the isolated²² small *Sph*I fragment of pSV₃gpt containing the *Ecogpt* gene and part of the SV40 early region⁹). The *gpt* probe was labelled *in vitro* by nick-translation²³, using [α -³²P]-labelled dATP and dCTP (Amersham, 3,000 Ci mmol⁻¹) to a specific activity of 2.5×10^8 c.p.m. μg^{-1} . During hybridization excess unlabelled pTCF plasmid was present as competitor. Hybridized and washed filters (Millipore, HA nitrocellulose, 0.45 μm pore size) were exposed to X-ray film (Fuji Rx) at -70°C with intensifying screen (Fuji). Seven hybridizing *E. coli* clones were picked from the master filter, rescreened with the same probe and grown in 1 l culture medium (L-broth + ampicillin, 50 $\mu\text{g ml}^{-1}$). Cosmid DNAs were isolated according to ref. 20. All cosmids transfected into 43-3B as described above using 1–5 μg of cosmid DNA gave MPA^r colonies; only cosmid 43-34 (see Figs 2–4) induced mitomycin C-resistance as well.



the right of pSV₃gptH and extends at least in part beyond the overlap with cosmid 43-23 (Fig. 4). This is consistent with the finding that a tertiary transformant, 43IIIMM-C, has retained only the single pSV₃gptH copy and the same fragment of human DNA (Fig. 2a, e). A more precise localization of the gene was obtained by digesting 43-34 cosmid DNA with various restriction endonucleases and scoring for the intactness of the gene by transfection into the 43-3B cells. Of the enzymes tested, only *SalI* and *ClaI* left the gene intact; other endonucleases (*Bam*HI, *Bgl*II, *Eco*RI, *Hind*III, *Kpn*I, *Nun*II, *Pst*I, *Pvu*II, *Sma*I, *Sst*I, *Xho*I) largely or completely abolished MM-C resistance. From these data we conclude that at least sites X₁, K₃, Bg₇, H₆ and B₂ are within essential parts of the DNA repair gene, leaving a minimal gene size of 7 kbp (Fig. 4, lower part).

Fig. 2 Fifteen μ g of high-molecular weight DNA of 43IUUV (primary transformant), 43IIIMM-C (secondary transformant) and 43IIIMM-C (tertiary transformant) were digested with *Pst*I. Three μ g of cosmids 43-12, -31 and -34 (see Fig. 4) were cleaved with *Sal*I to liberate the insert (Fig. 4; ref. 21) and further digested with *Pst*I. The DNAs were electrophoresed and hybridized with the probes as indicated. Arrows indicate non-equimolar bands, hybridizing with the *Ecogpt* probe, that are derived from contaminating free pSV₃gptH molecules in the cosmid DNA preparation. As these bands were only found in preparations of cosmids harbouring the tandem copies of pSV₃gptH, we presume they have arisen during growth of the bacteria for preparation of cosmid DNA as a result of a recombination event, which might be due to the presence of three to four pBR and at least two *Ecogpt* sequences in these cosmid molecules.

Methods: DNAs were prepared as detailed elsewhere^{17,21}. Genomic DNAs were ethanol-precipitated before layering onto a 0.5% agarose gel and electrophoresis. After transfer to a nitrocellulose filter²⁴, one set of blots was hybridized with the ³²P-labelled *Ecogpt* probe (Fig. 1 legend), the other with a ³²P-labelled human *cot-1* DNA probe, ligated using T4 DNA ligase before nick-translation to increase its size. Nick-translation, hybridization, washing (0.3 $\times$ SSC at 65 $^{\circ}$ C) and autoradiography were as described in Fig. 1.

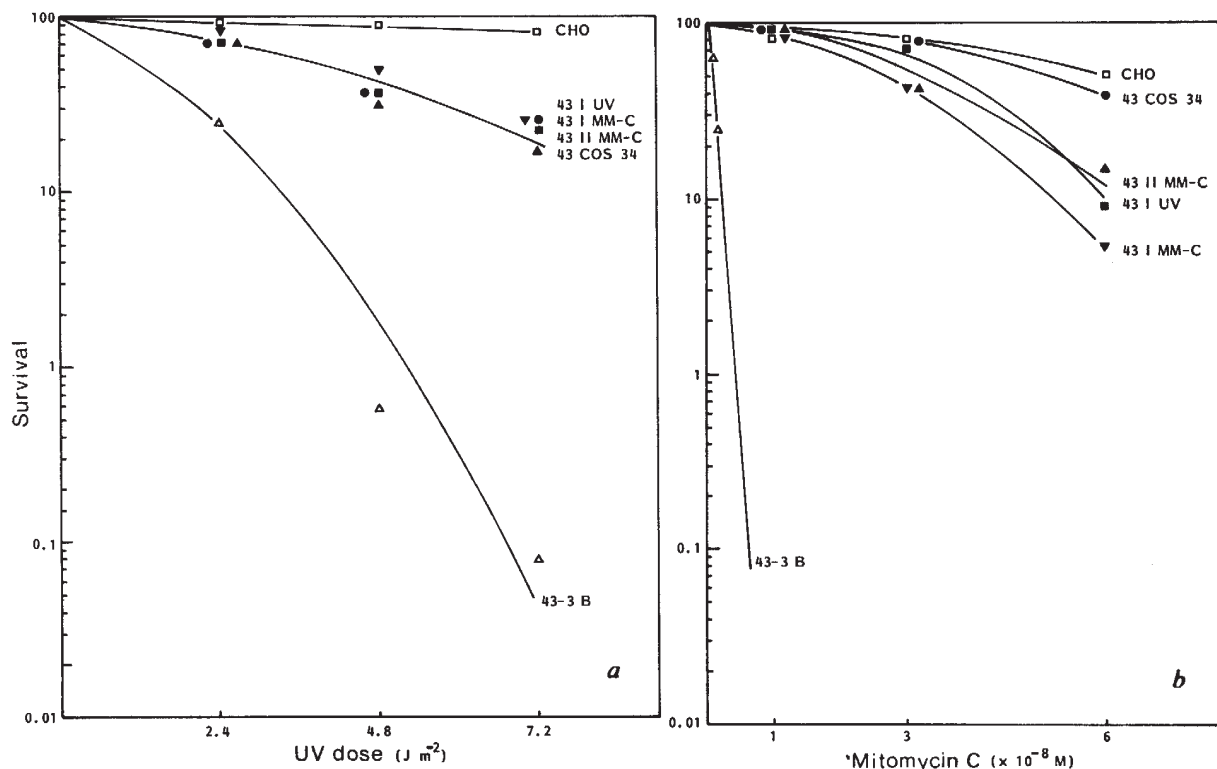
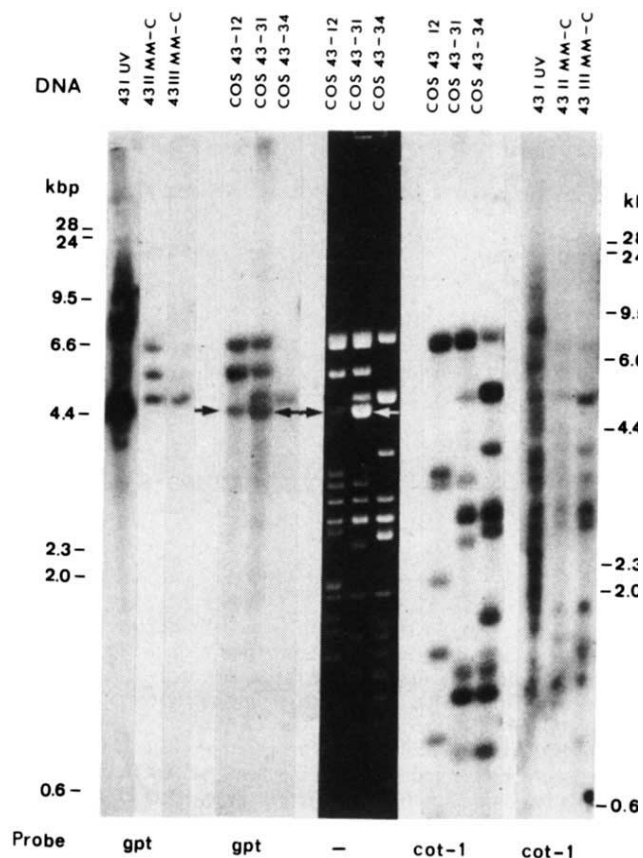


Fig. 3 UV (a) and mitomycin C (b) sensitivities of CHO ($\square$), 43-3B (Δ) and the primary (43IIIMM-C, 43IUUV), secondary (43IIIMM-C) and cosmid (43cos34) transformants (∇ , $\blacksquare$, $\blacktriangle$ and $\bullet$, respectively). Cells in normal growth medium were inoculated into 60 or 100 mm Petri dishes at densities of 2×10^2 – 10^4 cells per dish. One day after seeding the cells were either irradiated with 254 nm UV from a low-pressure mercury, germicidal lamp (Philips TUV lamp) at a dose rate of $0.6 J m^{-2} s^{-1}$, or grown in medium with mitomycin C. The cells were cultured for 7 days after which period the clones were fixed, stained and scored.

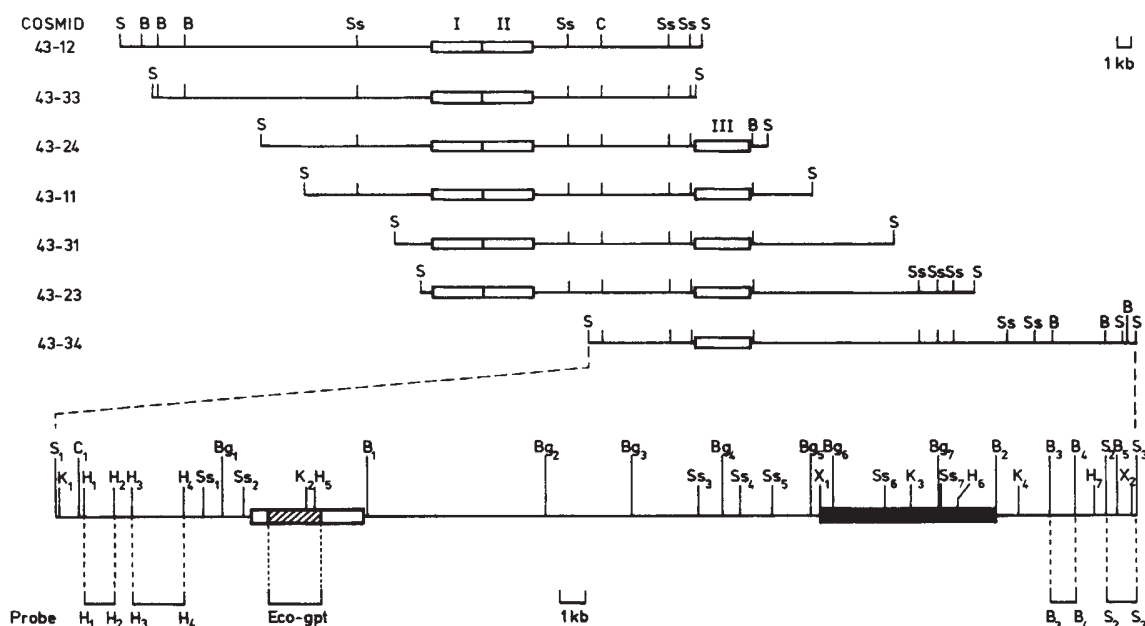


Fig. 4 Upper part, physical maps of seven cosmids covering the transfected *Ecogpt* region of 43IIMM-C. The pSV₃gptH copies, designated I–III, are indicated by open bars. Lower part, detailed physical map of cosmid 43-34. The pSV₃gptH copy is indicated as an open bar (the hatched region is the *Ecogpt* probe used). The minimal region covered by the repair gene *ERCC1* is drawn as a solid bar. Probes used for hybridization of 43-3B transformants are indicated separately. Restriction enzymes include: *Bam*HI (B), *Bgl*II (Bg), *Cla*I (C), *Hind*III (H), *Kpn*I (K), *Sal*I (S), *Sst*I (Ss) and *Xho*I (X).

Methods: Restriction enzyme maps are based on single, double and triple endonuclease digestion, on hybridization data with the *Ecogpt* and human *cot-1* probes and individual fragments isolated from low-melting point agarose gels²², and on subclones from insert regions. To obtain information on the location of the repair gene, 10 µg of cosmid 43-34 DNA were cleaved with various restriction endonucleases (see text) and co-transfected with pSV₃gptH to 43-3B cells. To prevent reconstitution of the gene by religation of the fragments in the cell after transfection, the restricted cosmid DNA was treated before transfection with alkaline phosphatase and, when appropriate, a molar excess of human DNA fragments with the same (cohesive) ends was added as carrier DNA. One µg of digested cosmid DNA was applied to each Petri dish, containing $0.4\text{--}1.0 \times 10^6$ 43-3B cells. Selection was based on MPA and mitomycin C-resistance simultaneously.

Consistent with this conclusion is the finding that the largely unique 1.0-kbp B₃–B₄ fragment (Fig. 4, lower part) was present in all of the 13 independent primary transformants checked, and some of them had retained the more distant 1.1-kbp S₂–S₃ fragment also. As expected, none of these transformants contained two unique *Hind*III fragments (H₁–H₂ and H₃–H₄; Fig. 4, lower part) from the other side of pSV₃gptH, that is, from a region unrelated to the repair gene (data not shown). Following the nomenclature proposed at the 17th Human Gene Mapping Conference¹⁵, we term the repair gene described here *ERCC1* (excision repair complementing defective repair in Chinese hamster cells).

Table 1 UV-induced unscheduled DNA synthesis of 43-3B transformants

Cell line	No. of grains
HeLa	37.0 ± 0.9
CHO	21.6 ± 0.8
43-3B	4.5 ± 0.2
43IUV*	17.6 ± 1.7
43IIMM-C	25.0 ± 0.7
43IIIMM-C	21.9 ± 0.5
43IIIMM-C	31.8 ± 0.6
43Cos34	31.5 ± 1.2

Two days after seeding the cells were exposed to UV light (16 J m^{-2}) and incubated for 2 h in F10 medium without TdR containing ³H-thymidine (10 µCi ml^{-1} ; specific activity 46 Ci mmol^{-1}) and 8% dialysed fetal calf serum. After fixation of the cells the preparations were processed for autoradiography (Kodak AR10 stripping film), exposed for 1 week at 4°C, developed, fixed and stained with Giemsa solution. The table shows the results of one typical experiment. For each cell line at least 50 nuclei were counted. Number of grains is given as mean ± s.e.m. per fixed square of non-S-phase nucleus.

* Note that about 20% of the 43IUV population did exhibit 43-3B unscheduled DNA synthesis levels; this might be due to segregation of the repair marker.

Other groups have reported the transfer and identification of human repair genes. In one case the transfer of human DNA conferring UV resistance to xeroderma pigmentosum cells of complementation group A has been claimed¹⁶ but this was not followed by molecular identification or cloning of the responsible gene. Extensive efforts to reproduce these results were unsuccessful in our hands (unpublished observations). Recently, a human repair gene was identified⁸ following gene transfer into a Chinese hamster mutant of the same complementation group as mutant 43-3B used in this study. The molecular cloning of the repair gene reported here makes possible an investigation of its function in repair, its relation to human repair-deficiency syndromes such as XP and general characteristics of mammalian repair processes.

We thank Drs G. C. Grosveld, A. de Klein, N. G. J. Jaspers and M. James for stimulating discussions and helpful suggestions; J. A. Bruyn, H. Heyboer and R. Kanaar for help in the experiments; and Mrs R. J. Boucke for typing the manuscript and Mr T. van Os for photography. This work was supported by Euratom contract no. BIO-E-404-NL and Fungo, Foundation of Medical Scientific Research in the Netherlands, contract 13–23–35.

Received 2 April; accepted 15 June 1984.

- Kraemer, K. H. in *Update: Dermatology in General Medicine* (eds Fitzpatrick, T. B., Eisen, A. Z., Wolff, K., Freedberg, I. M. & Austen, K. F.) 113–141 (McGraw-Hill, New York, 1983).
- Lehmann, A. R. & Karran, P. *Int. Rev. Cytol.* **72**, 101–146 (1981).
- Thompson, L. H., Busch, D. B., Brookman, K., Mooney, C. L. & Glaser, D. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3734–3737 (1981).
- Shiomi, T., Hieda-Shiomi, N. & Sato, K. *Somat. Cell Genet.* **8**, 329–345 (1980).
- de Weerd-Kastelein, E. A., Keijzer, W. & Bootsma, D. *Nature new Biol.* **238**, 80–83 (1972).
- Moshell, A. N. *et al.* in *Cellular Responses to DNA Damage* (eds Friedberg, E. C. & Bridges, B.) 209–213 (Liss, New York, 1983).
- Wood, R. D. & Burki, H. J. *Mutat. Res.* **95**, 505–514 (1982).
- Rubin, J. S., Joyner, A. L., Bernstein, A. & Whitmore, G. F. *Nature* **306**, 206–208 (1983).
- Mulligan, R. C. & Berg, P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2072–2076 (1981).
- Wigler, M. *et al.* *Cell* **16**, 777–785 (1979).
- Lowy, I. *et al.* *Cell* **22**, 817–823 (1980).
- Pellicer, A. *et al.* *Science* **209**, 1414–1422 (1980).
- Scangos, G. & Ruddle, F. H. *Gene* **14**, 1–10 (1981).

14. Hori, T.-A., Shiomi, T. & Sato, K. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5655-5659 (1983).
 15. Shows, T. B., McAlpine, P. J. & Miller, R. J. in *Proc. 7th Human Gene Mapping Conference* (eds Sparkes, R. S., Berg, K., Evans, H. J. & Klinger, H. P.) 340-390 (Karger, Basel, 1984).
 16. Takano, T., Noda, M. & Tamura, T. *Nature* **296**, 269-270 (1982).
 17. Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 429-439 (1977).
 18. Mulligan, R. E. & Berg, P. *Molec. cell. Biol.* **1**, 449-459 (1981).
 19. Graham, F. & Van der Eb, A. *Virology* **52**, 456-467 (1973).
 20. Grosveld, F. G., Dahl, H. H. M., de Boer, E. & Flavell, R. A. *Gene* **13**, 227-237 (1981).
 21. Grosveld, F. G. *et al. Nucleic Acids Res.* **10**, 6715-6732 (1982).
 22. Groffen, J. *et al. Cell* **36**, 93-99 (1984).
 23. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237-251 (1977).
 24. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).

Predicted structure of the sugar-binding site of the *lac* repressor

Clarence F. Sams, Nand K. Vyas, Florante A. Quiocho & Kathleen Shive Matthews

Department of Biochemistry, Rice University, Houston, Texas 77251, USA

The lactose repressor protein from *Escherichia coli* binds sugars, primarily galactosides, which modulate its interactions with operator DNA and thereby affect synthesis of the *lac* metabolic enzymes¹. The affinity of the repressor for operator DNA is decreased by binding inducer sugars and increased by binding anti-inducer sugars². Based on regions of the primary structure implicated by genetic methods to be involved in sugar binding³⁻⁵, amino acid sequence homology between L-arabinose-binding protein (ABP) and *lac* repressor has recently been reported⁶. The sugar-binding sites for these two proteins might be expected to have similar structural features, as both bind L-arabinose and D-galactose. The high resolution structure of ABP reported in the accompanying article⁷ provides complete definition of amino acids in the sugar-binding site. By identification of homologous residues in the *lac* repressor, we have now predicted the structure of the portion of the repressor sugar-binding site which accommodates the galactosyl moiety. This prediction provides the first potential view of the inducer/anti-inducer site in the repressor protein.

Although the X-ray structure analysis of the liganded form of ABP previously revealed several amino acid residues that are crucial for binding L-arabinose⁸, equivalent residues were not identified in *lac* repressor in the subsequent study of homologous amino acids⁶. In addition, homology was not examined in the regions containing Lys 10-Glu 14 and Asp 89-Asp 90 in ABP (Fig. 1). Of the three possible homologues to Lys 10-Glu 14 in the sugar-binding core domain^{9,10} of *lac* repressor, only the 84-88 amino acid region contains mutations which affect inducer binding^{4,11}; thus, this region was used in constructing the repressor binding site. Adjacent acidic residues similar to Asp 89-Asp 90 in ABP are present in repressor at amino acids 129-130 and 164-165. The 129-130 pair is adjacent to a residue at which mutation affects inducer binding of *lac* repressor protein^{4,11} and was therefore selected for substitution for repressor.

Figure 2 shows the ABP-binding site with *lac* repressor amino acid substitutions. The unreactive nature of the amino acids in the site readily explains the inability to generate chemical affinity labels for this site (of ~12 sugar compounds synthesized, none significantly affected inducer binding; K.S.M. *et al.*, unpublished data). Note that, contrary to the limited region for sugar interaction presumed on the basis of genetic studies⁶, amino acid residues throughout the ABP molecule participate in sugar binding; the amino acid residues that are involved in sugar interaction are derived from both structural domains of the protein⁷.

Several geometrical features that would be uniquely required for the repressor (as opposed to the ABP) sugar-binding site are produced as a consequence of the homologous amino acid substitutions made in the ABP site. The protein backbone of ABP is shown in Fig. 3 with disaccharides of galactose and glucose in the sugar-binding site. Substituents at the 1-position are limited to a β -orientation due to steric constraints, as an α -substituent will protrude almost perpendicularly into the protein structure. In contrast, a β -substituent falls parallel to and

ABP ...Lys 10-Glu 11...Glu 14...Asp 89-Asp 90...Thr 147...Arg 151...
 LAC ...Lys 84-Ser 85...Asp 88...Asp 129-Asp 130...Ser 193...Arg 197...
 ABP ...Met 204-Asn 205...Thr 208...Asn 232...Asp 235...
 LAC ...Ala 245-Asn 246...Met 249...Val 271...Asp 274...

Fig. 1 Homologies between the primary structures of *lac* repressor (LAC) and arabinose-binding protein (ABP) for residues in the ABP sugar site. The homologies past Thr 147 in ABP and Ser 193 in repressor are taken from Müller-Hill⁶. *Lac* repressor residues are in bold print.

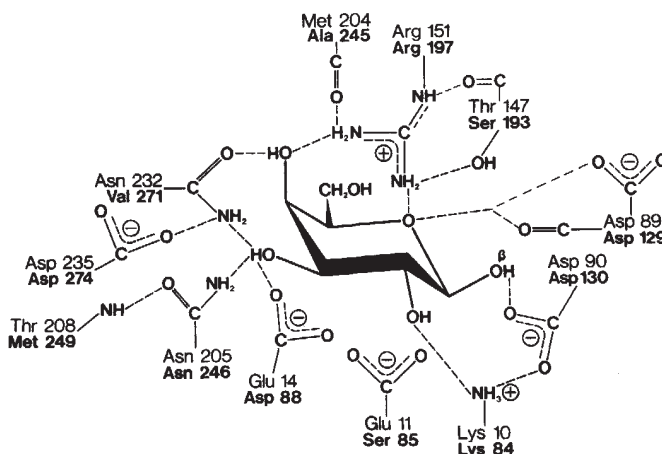


Fig. 2 Diagram of the sugar-binding site in arabinose-binding protein, with indicated amino acid substitutions in *lac* repressor sugar-binding site. Amino acids of the ABP site are indicated by normal lettering, while the *lac* repressor amino acids are indicated by bold lettering. The site has been drawn with the β -anomer of D-galactose present. Note that the ring oxygen interaction with Asp 89 (Asp 129 in *lac* repressor) is mediated through two H₂O molecules in ABP⁷.

within the cleft and thus fits easily into the structure. Consistent with this observation, the *lac* repressor binds α -galactosides^{2,12} weakly, if at all. In ABP, all substituents at the 1-position are precluded by the presence of the Thr 147 methyl group; however, substitution of Ser 193 into the site for repressor opens the 1-position for β -substitution, as required for this protein. Introduction of more bulky amino acids at position 193 by suppression of nonsense mutation has yielded mutant repressor proteins with decreased affinity for inducer¹¹. Further, in the region beyond the 1-position, where the substituents would interact with the protein structure, a hydrophobic contact with Leu 145 in ABP is possible; Ser 191 is the corresponding amino acid in *lac* repressor, with a leucine at position 189. Asp 90 in ABP (Asp 130 in repressor) could also make contacts in such a second site. The increased affinity of methyl, isopropyl and phenyl 1-substituted galactosides can be attributed to the formation of hydrophobic interactions with this area of the protein, while hydrogen bonding to the hydroxyls would facilitate 1,6-allolactose binding. Substitution at the 6-position of the sugar is possible in the postulated site; consistent with this, fucosides and arabinosides have been shown to interact with the *lac* repressor protein^{2,12}. Substituents or epimers at positions 2 or 3 would be unable to occupy the site due to severe steric hindrance; compounds with these characteristics do not bind to *lac* repressor^{2,12}. The changes in the ABP site to create a *lac* repressor site would weaken binding to arabinose⁷, but would increase galactose and particularly β -galactoside binding.

The *lac* repressor will bind both D-glucose and D-galactose, epimers which differ at the 4-position of the sugar and also in their effects on operator affinity (galactose is an inducer, while glucose is an anti-inducer). However, the affinities for these two hexoses differ by almost 1,000-fold. Glucose will not

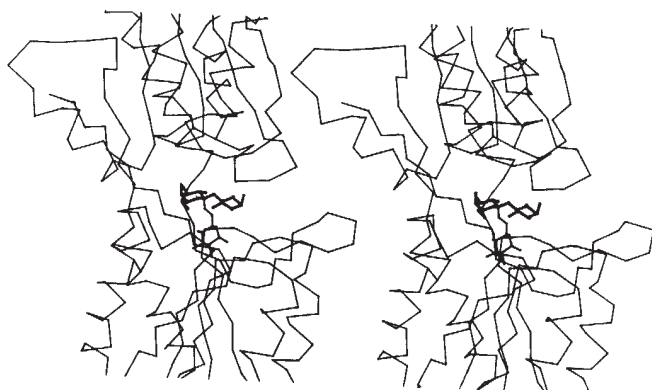


Fig. 3 Stereo view of the ABP binding site (thin lines) with lactose (4-O-β, D-galactopyranosyl-D-glucose; thick lines) and mellibiose (6-O-α, D-galactopyranosyl-D-glucose; medium lines). Only the C₄ chain of ABP is drawn. Note that only the β-substituted galactosides can be accommodated in the binding cleft. The coordinates of lactose²¹ and mellibiose²² derived from small molecule crystal structure analysis were used. The D-galactose pyranose ring of both disaccharides is superposed to the refined L-arabinose bound to ABP⁷.

fit into the postulated *lac* repressor site due to steric hindrance at the 4-position by amino acid side chains. Both allolactose, the natural inducer, and lactose, an anti-inducer, are disaccharides where the 1-substituent is glucose; thus, the presence of a second hexose binding site is suggested. Weak glucose binding may normally occur at the 1-β-substituent subsite rather than in the primary sugar-binding site of repressor.

In ABP, a conformational change which involves a decrease in the radius of gyration of the protein occurs in response to sugar binding¹³. These data have been interpreted to indicate that ligand binding results in a closing of the cleft containing the sugar-binding site^{8,13}. A conformational change in the core region of the repressor protein also occurs in response to inducer binding^{14–20}. The structural changes in the repressor which would be associated with cleft closing provide a potential mechanism for the observed alteration in operator DNA affinity concomitant with inducer binding. The relatively weak binding and opposite effects on operator affinity observed for anti-inducers may be postulated to result from their inability to elicit closing of the binding site. This predicted sugar-binding site in *lac* repressor provides a new and potentially useful view for designing studies of this important and largely unknown region of the repressor protein.

This work was supported by grants from the NIH (GM21371, GM22441), NSF (PCM82-14057) and the Robert A. Welch Foundation (C-581 and C-576).

Received 9 January; accepted 10 May 1984.

1. Miller, J. H. & Reznikoff, W. S. *The Operon* (Cold Spring Harbor Laboratory, New York, 1979).
2. Barkley, M. D., Riggs, A. D., Jobe, A. & Bourgeois, S. *Biochemistry* **14**, 1700–1712 (1975).
3. Pfahl, M., Stocker, C. & Gronenborn, B. *Genetics* **76**, 669–679 (1974).
4. Miller, J. H. *J. molec. Biol.* **131**, 249–258 (1979).
5. Müller-Hill, B., Gronenborn, B., Kania, J., Schlotmann, M. & Beyreuther, K. in *Nucleic Acid-Protein Recognition* (ed. Vogel, H. J.) 219–236 (Academic, New York, 1977).
6. Müller-Hill, B. *Nature* **302**, 163–164 (1983).
7. Quijcho, F. A. & Vyas, N. K. *Nature* **310**, 381–386 (1984).
8. Newcomer, M. E., Gilliland, G. L. & Quijcho, F. A. *J. biol. Chem.* **256**, 13213–13217 (1981).
9. Platt, T., Files, J. G. & Weber, K. *J. biol. Chem.* **248**, 110–121 (1973).
10. Matthews, K. S. *J. biol. Chem.* **254**, 3348–3353 (1979).
11. Miller, J. H. *et al. J. molec. Biol.* **131**, 191–222 (1979).
12. Müller-Hill, B., Rickenberg, H. V. & Wallenfels, K. *J. molec. Biol.* **10**, 303–318 (1964).
13. Newcomer, M. E., Lewis, B. A. & Quijcho, F. A. *J. biol. Chem.* **256**, 13218–13222 (1981).
14. Matthews, K. S. *Biochim. biophys. Acta* **359**, 334–340 (1974).
15. Laiken, S. L., Gross, C. A. & von Hippel, P. H. *J. molec. Biol.* **66**, 143–155 (1972).
16. Yang, D. S., Burgum, A. A. & Matthews, K. S. *Biochim. biophys. Acta* **493**, 24–36 (1977).
17. Alexander, M. E., Burgum, A. A., Noall, R. A., Shaw, M. D. & Matthews, K. S. *Biochim. biophys. Acta* **493**, 367–379 (1977).
18. O'Gorman, R. B. & Matthews, K. S. *J. biol. Chem.* **252**, 3565–3571 (1977).
19. Boschelli, F., Jarema, M. A. C. & Lu, P. *J. biol. Chem.* **256**, 11595–11599 (1981).
20. Jarema, M. A. C., Lu, P. & Miller, J. H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2707–2711 (1981).
21. Fries, D. C., Rao, S. T. & Sundaralingam, M. *Acta crystallogr.* **B27**, 994–1005 (1971).
22. Kanters, J. A., Roelofs, G., Doesburg, H. M. & Koops, T. *Acta crystallogr.* **B32**, 2830–2837 (1976).

Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250 °C

Robert H. White

Department of Biochemistry and Nutrition, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA

The upper temperature at which a living system can exist is limited by the hydrolytic breakdown rate of its chemical constituents. The peptide bonds of proteins, the phosphodiester and *N*-glycosyl bonds in RNA and DNA, and the pyrophosphate and *N*-glycosyl bonds in nucleotides such as ATP and NAD are among the more important bonds that will undergo hydrolysis. The decomposition of biomolecules via non-hydrolytic pathways such as decarboxylations and dehydrations may also be critical factors in determining this upper temperature limit. Baross and Deming¹ recently reported 'black smoker' bacteria, which they isolated from deep-sea hydrothermal vents, growing at 250 °C. Here I have attempted to establish the rates for the hydrolysis and/or decomposition of critical biomolecules to determine their ability to exist at this temperature. My results clearly indicate that if these organisms exist, and if their metabolic reactions occur in an aqueous environment, they could not survive at this temperature if they were composed of biomolecules such as proteins and nucleic acids, due to the very rapid rate of decomposition of such molecules.

The recorded maximum upper temperature for the growth of microorganisms has increased from 55 °C in 1903 to 93.5–95.5 °C in 1970², and the growth of organisms at 105 °C has recently been described³. However, the few data available on the temperature stability of biomolecules in neutral aqueous solutions at temperatures above 100 °C have dealt with the decomposition of hot aqueous solutions of amino acids in order to simulate the geochemical degradation of organic matter^{4,5}. Extrapolation of the limited data for hydrolysis or decomposition rates at temperatures of 100–250 °C is difficult because of the possible non-linearity of the free energy of activation over this wide a temperature range. To avoid this problem, and to obtain additional data for molecules for which no low-temperature data are available, the decomposition rates for various biomolecules in water were measured at 250 °C.

The data show (Table 1) that the peptide bonds in the alanine peptides studied have a hydrolysis half life ($t_{1/2}$) of ~7 min at 250 °C. However, not all peptides are as stable as the alanine peptides, as evidenced by the very rapid breakdown rate (<1 min) of the Ala-Asp and Glu-Ala peptides. For all peptides studied, except those containing aspartic acid, peptide bond hydrolysis was confirmed by the increase in the concentration of the expected free amino acid hydrolysis products with time. The rapid deamination of the aspartic acid released from those peptides containing this amino acid precluded its detection by amino acid analysis⁶.

A $t_{1/2}$ of 1.08 s for a 48,000-molecular weight (MW) protein was calculated using the $t_{1/2}$ for the hydrolysis of the (Ala)₃ peptide of 7.2 min. This value is based on the assumption that the average molecular weight of amino acids in proteins is 120 and that each of the peptide bonds in the protein will hydrolyse with the same $t_{1/2}$ as observed for (Ala)₃. Assuming that the hydrolysis of any peptide bond will destroy the protein, a 48,000-MW protein will be destroyed 400 times more rapidly than a single peptide bond and thus have a $t_{1/2}$ of 1.08 s. One would expect the true $t_{1/2}$ for a protein containing all the amino acids to be much shorter, however, due to the very rapid cleavage of the proteins at the aspartic acid residues⁷. This expectation was confirmed by the rapid hydrolysis observed for the Ala-Asp peptide. Attempts to measure the hydrolysis rate of intact proteins (lysozyme and bovine serum albumin) at 250 °C by SDS-gel electrophoresis⁸ were unsuccessful due to the very rapid breakdown of these proteins. No Coomassie blue-stainable bands

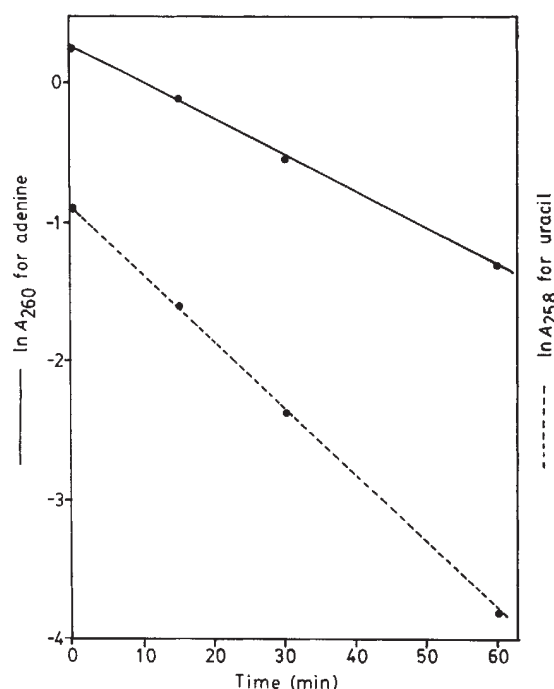


Fig. 1 Kinetic data for the decomposition of adenine and uracil at 250 °C in 0.1 M potassium phosphate buffer at pH 7.00. Data were determined as described in Table 1 legend.

were detected for the intact protein nor for peptide fragments after heating the protein solution in an oil bath at 250 °C for periods of 3 min or more.

A second mechanism for the breakdown of proteins, recently investigated by Steinberg and Bada⁹, is the cyclization of the last two amino acids in the amino-terminal end of the protein to form a diketopiperazine and the protein minus these two amino acids. If this cyclization is a very facile process, it could also lead to the rapid breakdown of the proteins at high temperature. Modification of the terminal amino group (acetylation being one method) would block this process and would therefore be a simple mechanism by which an organism could prevent this type of decomposition. However, as the *N*-acetyl (Ala)₄ peptide was hydrolysed at about the same rate as (Ala)₄ (Table 1), this does not seem to be an important factor in the hydrolytic breakdown of this peptide at 250 °C.

Many of the amino acids released from proteins decompose with $t_{1/2}$ values of 1–5 h (Table 1). As these rates were measured in mixtures containing many of the amino acids, the difference between the value reported for alanine decomposition at 250 °C (380 min) and that calculated from the Arrhenius equation derived by Vallentyne⁵ from Abelson's⁴ data (904 min) can be explained by the increased rate of decomposition of the amino acid in a mixture⁵. An important exception to the decomposition of amino acids with $t_{1/2}$ values of 1–5 h is aspartic acid, which decomposes far too rapidly to be measured by the methods reported here. Extrapolation of the data of Bada and Miller⁶ for the deamination of aspartic acid to 250 °C yields a $t_{1/2}$ of only 7.5 s. In addition to the rapid decomposition of the amino acids, we would expect protein-bound amino acids to racemize readily at 250 °C. This racemization has been observed at temperatures much less than 250 °C for both free amino acids¹⁰ and amino acids bound either as proteins¹¹ or as peptides derived from proteins^{12,13}.

Nucleic acids also undergo hydrolytic breakdown of the primary structure followed by the decomposition of each of the components. No attempt was made to measure the hydrolysis rate of a high molecular weight nucleic acid because the rapid hydrolysis of bis-*p*-nitrophenyl phosphate, which served as a model for the nucleic acid phosphodiester group, suggested an

Table 1 Half lives for the hydrolysis and decomposition of various biomolecules

Peptide bonds	$t_{1/2}$	Phosphodiester bonds	$t_{1/2}$
(Ala) ₂	5.8 min	bis- <i>p</i> -Nitrophenyl phosphate	2.3 min
(Ala) ₃	7.2 min	Diethyl phosphate	420 min
(Ala) ₄	5–11 min	<i>E. coli</i> DNA (2.5 × 10 ⁹ MW)	19.3 μs §
<i>N</i> -acetyl(Ala) ₄	11.8 min		
Ala-Asp	< 1 min		
Glu-Ala	< 1 min		
48,000 MW protein	< 1.08 s*	Glycosidic bonds	
Amino acids derived from lysozyme		Adenosine	1.3 min
Ala	380 min	ATP	< 1 s
Lys	105 min	Carbon–nitrogen bond in purines and pyrimidines	
Pro	120 min	Uracil	15 min
Ilu	65 min	Thymine	57 min
Leu	100 min	Adenine	27 min
Val	230 min	Cytosine	29 min
Tyr	100 min		
Free amino acids			
Asp	7.5 s †		
Tyr	278 min		
Ser	5.3 min ‡		
Thr	0.75 min ‡		

* Calculated value assuming the rate of hydrolysis of each peptide bond in a protein to be the same as that observed for (Ala)₃.

† Calculated from the equation, $\log k_{\text{deam.}} (\text{s}^{-1}) (pH 7.0) = 14.35 - 8047.5/T$ given by Bada and Miller⁶ for the deamination of aspartic acid.

‡ Calculated from the data of Vallentyne⁵ for the amino acids dissolved in water⁵.

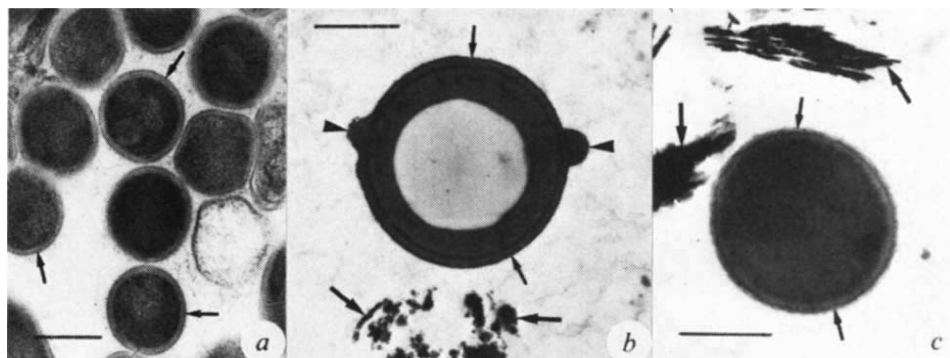
§ Calculated value assuming the rate of hydrolysis of each phosphodiester bond in DNA to be the same as that observed for bis-*p*-nitrophenyl phosphate.

|| The specific bond decomposition (glycosidic, phosphate ester or phosphate anhydride) responsible for the destruction of the ATP was not determined.

Methods: Solutions (200 ml of 1 mM) of the indicated biomolecules were prepared in 0.1 M potassium phosphate buffer at pH 7.00 and degassed for 15 min with oxygen-free nitrogen. Three 100-μl portions of this solution were transferred anaerobically to a 6 × 50 mm Kimax test tube which had been previously flushed with nitrogen. Each sample tube was then sealed by melting it just above the liquid level and mounted in a small wire holder so that the samples could be immersed and withdrawn from a silicone oil bath heated at 250 ± 2 °C. Generally, a set of samples was heated simultaneously and individual samples were removed as a function of time. The exact time interval was based on the results of test hydrolyses and allowed a minimum of four to five samples to be assayed, with the longest time point corresponding to at least 86% sample destruction. After the samples had cooled to room temperature, the concentration of the starting material and/or the product was assayed as described below. The $t_{1/2}$ was obtained by plotting the log of the sample concentration or absorbance against time. The slope of the resulting line divided into 0.693 gave the reported $t_{1/2}$. No change in the pH of the solutions was observed. Based on published data for the pH of phosphate buffers at 25–150 °C, the pH of the solutions used here were expected to be within 0.5 pH units of 7.0 at 250 °C^{29,30}. When heated, the contents of the sealed tubes reached 98.7% of their final temperature within ~1 min. This was measured by a thermocouple inserted into an open sample tube containing silicone oil instead of buffer. Peptide hydrolysis and amino acid decomposition were established by measuring the reduction in concentration of the peptide or amino acid with time using a Beckman Model 121 automatic amino acid analyser. Hydrolysis of adenosine was established by both HPLC and gas chromatography. HPLC allowed for the separation of adenosine from adenine which, in turn, allowed for the measurement of both the destruction of the adenosine and the production of adenine³¹. Gas chromatography allowed for the quantitation of the adenosine after conversion into its trimethylsilyl (TMS) derivative³². Decompositions of adenine, cytosine, uracil, thymine and tyrosine were determined by measuring the decrease in absorbance at their respective λ_{max} of 260.5, 267, 259.5, 264.5 and 280 nm. Hydrolysis of bis-*p*-nitrophenyl phosphate was measured by the increase in absorption at 410 nm caused by the generation of the *p*-nitrophenyl anion. Hydrolysis of diethyl phosphate was measured by determining the concentration of unreacted material as a function of time. This was done by extracting the acidified samples with ethyl acetate and quantitating the recovered diethyl phosphate by gas chromatography of the monoTMS derivative. The derivative was prepared by reacting the diethyl phosphate with a mixture of pyridine, hexamethyldisilazane and trimethylsilyl chloride (9:3:1)³³. ATP concentration was measured using the luciferin–luciferase reaction³⁴.

extremely short half life for DNA. For example, using the observed $t_{1/2}$ for the hydrolysis of bis-*p*-nitrophenyl phosphate of 2.3 min and assuming that each phosphodiester bond in the *Escherichia coli* chromosome hydrolyses at the same rate, a DNA molecule the size of an *E. coli* chromosome (2.5 × 10⁹ MW), containing 7.14 × 10⁶ bases would be split with a $t_{1/2}$ of 19.3 μs [(2.3 min/7.14 × 10⁶) × (60 s/min)]. Even using the data for the much more stable diethyl phosphate ($t_{1/2}$ 420 min), one half of the chromosome would be split at one of the phosphodiester bonds in 3.5 ms. The rapid hydrolysis observed for the

Fig. 2 Transmission electron micrographs of ultra-thin sections of *Methanobacterium formicicum* before (a) and after (b, c) heating for 30 min at 250 °C. Small arrows, cell walls; triangles, bud scars. Scale bars: 0.25 μm (a), 1.0 μm (b) and 0.5 μm (c). The samples contained 80 mg wet weight of the packed cells suspended in 1 ml of the phosphate buffer. Cell suspensions were completely clear after 1 min of heating at 250 °C, indicating complete solubilization of the original cellular structure. On cooling, a precipitate developed which could be removed from the buffer by centrifugation. Small pieces (~1 mm²) of the resulting pellet and of a control pellet from non-heated bacteria were immobilized in 1% agar and simultaneously fixed in glutaraldehyde and osmium tetroxide by the method of Franke *et al.*³⁵. Specimens were washed briefly in cold distilled water, soaked for 8 h in aqueous 1% uranyl acetate at 4 °C, dehydrated in an ethanol and acetone series, and embedded in a mixture of Epon and Araldite³⁶. Thin sections were collected on Formvar-coated copper, 100-mesh grids and stained in aqueous uranyl acetate³⁷ and lead citrate³⁸. Specimens were examined with a Philips EM 300 electron microscope operated at 60 kV.



N-glycosyl bond in adenosine ($t_{1/2}$ 1.3 min) is consistent with these bonds being the least stable portion of the DNA molecule. The released purine and pyrimidine bases also readily decompose at 250 °C with $t_{1/2}$ values of <1 h (Table 1, Fig. 1). The stability of the phosphodiester bonds observed in DNA would not, however, be expected in RNA, as the formation of 2',3'-cyclic phosphodiesters would facilitate the hydrolysis¹⁴. This has in fact been observed in the low-temperature (80–100 °C) hydrolysis of RNA^{15,16} and DNA¹⁶.

The concentration of ATP, the major energy transducing molecule in cells, decreased three orders of magnitude after the sample had been heated for 30 s, a decomposition rate too rapid to be measured accurately by the methods reported here. As the very specific luciferin-luciferase assay was used to determine the ATP levels, it was impossible to determine which portion of the molecule was initially destroyed because the presence of a cleaved bond in the ATP would indicate a reduction in the ATP concentration. Chelation of the ATP with a metal ion, which might be expected to increase the stability of the ATP at high temperatures, has, in fact, been found to enhance the rate of decomposition^{17–19}. The metal-ion catalysed hydrolytic cleavage of peptide bonds at neutral pH has also been reported²⁰.

The above rates of decomposition at 250 °C can be better appreciated when they are compared with either the calculated or observed rates for the decomposition of biomolecules around 100 °C, the present maximum temperature for bacterial growth. Thus, transfer RNAs still retain 27% of their biologically active structures after 10 h at 90 °C. RNAs and DNAs of 10,000 MW have $t_{1/2}$ values¹⁶ of 5–15 min, respectively, at 100 °C. Based on the data of Tetras and Lowenstein¹⁷, ATP at pH 5.07 has a $t_{1/2}$ of 4.7 h at 80 °C, while peptide bonds have $t_{1/2}$ values of ~1.7 h at 130 °C⁹. These rates of decomposition are small enough to conceive of a cell growing at 100 °C, as has in fact, been observed³.

The data reported here represent decomposition rates at 39.2 atm, the vapour pressure of water at 250 °C²¹. As this pressure is only 14.8% of that used by Baross and Deming¹ in their growth studies of the 'black smoker' bacteria, the effect of the higher pressure on decomposition rates was evaluated. Published data on a wide range of chemical reactions clearly show that the effect of these relatively low pressures on decomposition rates is minimal and would, in general, lead to an increase in hydrolysis rates at higher pressures²². This increase has been observed for the hydrolysis of several biochemically related compounds^{23–26}. Therefore, increased pressure would not be expected to increase the stability of the biomolecules.

One of the advantages of the present experimental method is that one can observe visually the physical state of the sample through the glass tubes as the solutions are being heated—cloudy solutions indicate the presence of insoluble material and clear solutions indicate complete solubilization. Thus, suspensions of bovine sperm, *E. coli* and *Methanobacterium formicicum* cells

start out as cloudy solutions but become clear after a few minutes at 250 °C. These clear solutions then become cloudy on cooling and produce structures which have the appearance of cells (Fig. 2). Although the chemistry of these structures was not evaluated, they are clearly non-living material related to coacercates and microspheres which form when hot solutions of proteins and nucleic acids are cooled²⁷. Even in the absence of bacteria, structures similar to those shown in Fig. 2 have recently been produced with the same growth medium, temperature and pressure apparatus used by Baross and Deming²⁸. It is more likely that structures such as these, and not bacterial cells, were observed by Baross and Deming.

I thank Dr Dan Dylewski for obtaining the electron micrographs, Lori Failla for performing the ATP assays, Cindy Knudson for her help in performing some of the kinetic assays, Dr James Ferry for supplying the *M. formicicum* cells, and Linda D. White for editing the manuscript. This work was funded in part by NSF grant PCM-8217072 and grant J-38 funded by the Jeffress Foundation.

Received 7 February; accepted 23 May 1984.

- Baross, J. A. & Deming, J. W. *Nature* **303**, 423–426 (1983).
- Brook, T. D. *Thermophilic Microorganisms and Life at High Temperatures* (Springer, New York, 1978).
- Stitter, K. O. *Nature* **300**, 258–260 (1982).
- Abelson, P. H. *Scient. Am.* **195**, 83–92 (1956).
- Valentyne, J. R. *Geochim. cosmochim. Acta* **28**, 157–188 (1964).
- Bada, J. L. & Miller, S. L. *J. Am. chem. Soc.* **92**, 2774–2782 (1970).
- Schultz, J., Allison, H. & Grice, M. *Biochemistry* **1**, 694–698 (1962).
- Weber, K. & Osborn, M. *J. biol. Chem.* **244**, 4406–4412 (1969).
- Steinberg, S. M. & Bada, J. L. *J. org. Chem.* **48**, 2295–2298 (1983).
- Bada, J. L. *Adv. Chem. Ser.* (No. 106), 309–331 (1971).
- Liardon, R. & Hurrell, R. F. *J. agric. Food Chem.* **31**, 432–437 (1983).
- Smith, G. G. & de Sol, B. S. *Science* **207**, 765–767 (1980).
- Steinberg, S. M. & Bada, J. L. *Science* **213**, 544–545 (1981).
- Brown, D. M. & Todd, A. R. *J. chem. Soc.* **1952**, 52–58 (1952).
- Lindahl, T. *J. biol. Chem.* **242**, 1970–1973 (1967).
- Eigner, J., Boedtker, H. & Michaels, G. *Biochim. biophys. Acta* **51**, 165–168 (1961).
- Tetras, M. & Lowenstein, J. M. *Biochemistry* **2**, 350–357 (1963).
- Sigel, H. & Amsler, P. E. *J. Am. chem. Soc.* **98**, 7390–7400 (1976).
- Ramirez, F., Marecek, J. F. & Szamosi, J. *J. org. Chem.* **45**, 4748–4752 (1980).
- Long, D. A., Truscott, T. G., Cronin, J. R. & Lee, R. G. *Trans. Faraday Soc.* **67**, 1094–1103 (1971).
- Keyes, F. G. in *International Critical Tables Vol 3* (ed. Washburn, E. W.) 233 (McGraw-Hill, New York, 1928).
- Gonikberg, M. G. *Chemical Equilibria and Reaction Rates at High Pressures* (National Science Foundation, Washington, DC, 1963).
- Laidler, K. J. & Chen, D. *Trans. Faraday Soc.* **54**, 1026–1033 (1958).
- David, H. G., Hamann, S. D. & Lake, S. J. *Aust. J. Chem.* **8**, 285–288 (1955).
- Williams, E. G., Perrin, M. W. & Gibson, R. O. *Proc. R. Soc. A* **154**, 684–703 (1936).
- LeNoble, W. J., Srivastava, S., Breslow, R. & Trainor, G. *J. Am. chem. Soc.* **105**, 2745–2748 (1983).
- Fox, S. W. *The Origins of Prebiological Systems and of Their Molecular Matrices* (Academic, New York, 1965).
- Trent, J. D., Chastain, R. A. & Yayanos, A. A. *Nature* **307**, 737–740 (1984).
- Bates, R. G., Pinching, G. D. & Smith, E. R. *Bur. Stand. J. Res.* **45**, 418–429 (1950).
- Stene, S. *Recl. Trav. chim. Pays-Bas Belg.* **49**, 1133–1145 (1930).
- Ramos, D. L. & Schoffstall, A. M. *J. Chromatogr.* **261**, 83–93 (1983).
- Sasaki, Y. & Hashizume, T. *Analyt. Biochem.* **16**, 1–19 (1966).
- Sweeley, C. C., Bentley, R., Makita, M. & Wells, W. W. *J. Am. chem. Soc.* **85**, 2497–2507 (1963).
- Hammerstedt, R. H. *Analyt. Biochem.* **52**, 449–455 (1973).
- Franke, W. W., Krien, S. & Brown, R. M. *Jr. Histochemie* **19**, 162–164 (1969).
- Poolswat, S. S. *Proc. EMSA* **31**, 364–365 (1973).
- Watson, M. L. *J. biophys. biochem. Cytol.* **4**, 475–478 (1958).
- Venable, J. H. & Coggeshall, R. *J. Cell Biol.* **25**, 407–408 (1965).

The impossibility of inductive probability

POPPER and Miller are right! If inductive support exists (a conjecture Popper and Miller doubt is true), it cannot be probabilistic support. In their note¹ they claim to prove this thesis.

Let h = all emeralds are green, e = all observed emeralds are green and f = all unobserved emeralds are green. Popper and Miller point out that $h = (h \vee e)(h \leftarrow e)$, where $\vee$ signifies or and $\leftarrow$ signifies if. This conjunction 'factors' h into an ampliative component, $h \leftarrow e$, not entailed by e and a non-ampliative component, $h \vee e$, entailed by e .

Other such factorizations of h are possible. Jeffrey (see below) factors h into f and e ; and $h = f(f \rightarrow e)$ as well. The factorization favoured by Popper and Miller satisfies conditions (i) and (ii) of their note. This is why they favour it. Jeffrey's factorization into f and e violates conditions (i) and (ii), but the factorization into f and $f \rightarrow e$, which shares a common ampliative component with Jeffrey's factorization, satisfies the two conditions.

Popper and Miller show that the probabilistic support for the ampliative component cannot be positive. If inductive support is probabilistic support, the ampliative component cannot possibly receive positive inductive support—which is absurd.

Jeffrey points out that the inductive support for f (the ampliative of both fe and $f(f \rightarrow e)$) can be positive. If we use the factorization into f and $f \rightarrow e$, we have a factorization satisfying conditions (i) and (ii) where the ampliative component may receive positive probabilistic support.

This result does not undermine the Popper–Miller conclusion. On the contrary, it completes a missing step in their argument.

Two logically distinct hypotheses which are, nonetheless, equivalent given the total evidence, ought to receive the same inductive support from that evidence. The inductive support, relative to e , for h , f and $h \leftarrow e$ should all be the same since the truth of e entails that they are all equivalent. Hence, if we can show that there is a factorization relative to which the ampliative component cannot possibly receive positive probabilistic support, then either probabilistic support satisfies the elementary equivalence condition just cited so that no factorization allows the

ampliative component to have positive probabilistic support, or probabilistic support violates the equivalence condition. Probabilistic support violates the equivalence condition; but in either case, probabilistic support cannot be inductive support.

This reasoning does not license the further conclusion advanced by Popper and Miller that probabilistic support is purely deductive. As we have seen, probabilistic support speaks with many tongues.

The result which Popper and Miller (with help from Jeffrey) establish reminds us that Bayesians overreach themselves when they use measures of probabilistic support as indices of inductive support.

ISAAC LEVI

*Department of Philosophy,
Philosophy Hall,
Columbia University,
New York, New York 10027, USA*

1. Popper, K. & Miller, D. *Nature* 302, 687–688 (1983).

IN their last paragraph Popper and Miller¹ assert what I had thought Popper denied: 'There is such a thing as probabilistic support' (call it ' s '), that is,

$$s(h, e) = p(h, e) - p(h) \quad (1)$$

which can be positive (that is, support), negative (countersupport) or null (irrelevance). That clarifies matters.

As they point out, any hypothesis, h , can be expanded relative to any statement, e , as a conjunction

$$h = (h \leftarrow e)(h \vee e) \quad (2)$$

where the second factor is entailed by e , and where the factors are respectively unsupported and uncountersupported by e :

$$s(h \leftarrow e, e) \leq 0 \leq s(h \vee e, e) \quad (3)$$

(Except in trivial cases, both inequalities are strict, so that e supports the second

factor, and countersupports the first.) The support e gives h lies between the supports it gives the separate factors, for it is their sum:

$$s(h, e) = s(h \leftarrow e, e) + s(h \vee e, e) \quad (4)$$

So far, so good. But at the end they proclaim: "This result is completely devastating to the inductive interpretation of probability. All probabilistic support is purely deductive [because] that part of a hypothesis that is not deductively entailed by the evidence is always strongly counter-supported by the evidence..."

I say the reason given is specious, being based on this thought: The part of h that goes beyond e must be the weakest truth function of h and e which, conjoined with $h \vee e$, yields h . To see that this is false, consider the familiar case where h is a universal generalization to which all observations have conformed so far, for example, h = all are green, e = all we have seen are green. Here, the part of h that goes beyond e is clearly:

f = so are the rest

and $h \leftarrow e$ is another matter altogether: since $h = ef$, we have $h \leftarrow e = ef \leftarrow e = f \leftarrow e \neq f$ so that $h \leftarrow e$ is not f but $f \leftarrow e$, that is, if all we have seen are green, so are the rest. And the relevant factoring is not equation (2) but $h = fe$, where f is no truth function of h and e .

Observe that where h implies e , e must p -support h (unless $p(h) = 0$ or $p(e) = 1$), and so, normally, 'all we have seen are green' p -supports 'all are green'. Inductive support is probabilistic where it exists, for example, where 'all we have seen are green' increases the probability of 'so are the rest'. (It is $f \leftarrow e$, not f , that has to be countersupported by e .) But whether p -support is inductive or not depends on p , since, for example, $s(f, e) < 0 < s(h, e)$ when in place of 'green' above stands 'grue' (*viz.*, green if already observed, and otherwise blue). Here, indeed, e supports $h (=ef)$ only because it supports itself. But this is not always so.

RICHARD C. JEFFREY

*Department of Philosophy,
Princeton University,
Princeton, New Jersey 08544, USA*

1. Popper, K. & Miller, D. *Nature* 302, 687–688 (1983).

WHENEVER we assume that words probably mean roughly what they have meant in the past, and in particular whenever we write letters to *Nature*, we are relying on probabilistic induction^{1,2}. Therefore, when Popper and Miller³ claim to prove the impossibility of inductive probability this argument needs to be watertight if it is to be believed. Their argument falls short of achieving the potentially watertight (but unachievable) form stated in my next paragraph.

Let h be any hypothesis and e any event such that $p(e) \neq 1$ and $p(h, e) \neq 1$, and suppose that h can always be written in the form $h = ab$ where (1) e deductively implies b , (2) e probabilistically undermines a , that is $p(a, e) < p(a)$, and (3) a and b are probabilistically independent. All this is to be true even if $p(h, e) > p(h)$. This situation would I think be "completely devastating to the inductive interpretation of the calculus of probability". to quote Popper and Miller³.

Popper and Miller in effect define a as $h \leftarrow e$ and b as $h \vee e$. Then all conditions are satisfied, except that their condition (3) is replaced by the weaker condition: (3a) a and b are probabilistically independent when e is given (they are also independent given $\neg e$). But, apart from the uninteresting cases in which $p(h, e) = 1$ or $p(h, \neg e) = 1$ or $p(e) = 1$ or $p(\neg e) = 1$ (where $\neg$ signifies not), the independence of a and b is impossible, that is, $p(a)p(b) \neq p(ab)$ which is $p(h)$. This can be proved by elementary algebra (the proof is available on request). Because I believe that inductive probability cannot be refuted I predict that Popper and Miller will not be able to redefine a and b to satisfy the conditions of my second paragraph, which include condition (3) instead of only (3a). Thus, in accordance with Popper's standard and justifiable requirements, I have stuck my neck out and have handed him an axe.

After reading the text above, David Miller asked me in what way I "understand the splitting of a hypothesis into an established and an ampliative part"; that is to say, "how are we to specify which part of a hypothesis goes beyond the available evidence e ?" My definition would be the set H of those deductions from h that are not deductions from e . This definition is not the same as $\neg e \vee h$, or $h \leftarrow e$, which was the definition proposed by Popper and Miller. I think my definition has more intuitive appeal, and if it is used, Popper and Miller's argument against the possibility of induction would not apply.

Miller has asked further why I think that the part of h that goes beyond e should be statistically independent of e .

The answer is that I do not by any means think this: I raised the issue of the probabilistic independence of $h \vee e$ and $h \leftarrow e$ because Popper and Miller stated in their original letter³ that "given e , the factors of h are probabilistically independent" so I assumed that this was intended to be a link in their chain of reasoning. If it was not so intended, then the second and third paragraphs of the present letter can be ignored. Moreover, if my definition of the ampliative part, H , of h is accepted, then this issue of independence would again evaporate.

Miller has pointed out in correspondence that H is not a proposition, for it contains h , but does not contain $h \vee e$ even though $h \vee e$ is a deduction from H . He has pointed out further that H contains some statements, such as $h \leftarrow e$, that are undermined by e . These are interesting observations, but I still think that H , rather than $h \leftarrow e$, provides the natural meaning for the part of h that goes beyond e . Probabilistic induction might be complicated but it cannot be impossible for the reason given in the first sentence of this letter.

Note added in proof: If inductive support does not mean probabilistic support, I do not know what it means.

I. J. GOOD

Department of Statistics,
Virginia Polytechnic Institute and State
University,
Blacksburg, Virginia 24061, USA

1. Black, M. in *The Encyclopedia of Philosophy* Vol. 4, 168-181 (Macmillan, New York and The Free Press, Glencoe, 1967).
2. Good, I. J. *J. Statist. Comput. Simul.* 13, 154 (1981).
3. Popper, K. & Miller, D. *Nature* 302, 687-688 (1983).

POPPER AND MILLER REPLY—As Levi says, "probabilistic support speaks with many tongues"; in other words, between he , the strongest, and $h \leftarrow e$, the weakest statement to entail just h in the presence of e , there are many others, and each may have a different degree of probabilistic support given e . Let x be one of these statements, so that

$$he \vdash x \vdash h \leftarrow e$$

(where $\vdash$ signifies entailment). Then it follows simply from Jeffrey's equation (4) that

$$s(x, e) = s(x \vee e, e) + s(h \leftarrow e, e)$$

The first of the terms on the right is greater than or equal to zero, but purely deductive. So only the second term, which is never

positive, can be said to be non-deductive (and so perhaps inductive). Inductive support (if any) is therefore at most zero: all non-deductive support is countersupport. None of our three correspondents has said a word against this thesis: Levi seems to agree; Jeffrey's f is a special case of our x ; Good's letter begins and ends with a declaration of faith in probabilistic induction. In the rest of his letter (as he says, his second and third paragraphs can be disregarded) he proposes (like Jeffrey) a different factorization from ours. But the proof of our thesis does not depend on our factorization, only on the fact that $s(h \vee e, e)$ or $s(x \vee e, e)$ is purely deductive support; whilst $s(x \leftarrow e, e) = s(h \leftarrow e, e)$ is the only support without a deductive part that is obviously redundant in the presence of e .

Thus all probabilistic support that is not countersupport is purely deductive.

We wish to withdraw as incorrect the suggestion near the end of our letter of 21 April 1983 that $s(h \leftarrow e, e)$ invariably decreases as the content of e increases. Our main thesis is not affected by this correction.

KARL POPPER
DAVID MILLER

Fallowfield, Manor Close,
Penn, High Wycombe,
Buckinghamshire HP10 8HZ, UK
and
Department of Philosophy,
University of Warwick,
Coventry CV4 7AL, UK

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.

Concepts of change

A. Hallam

Catastrophes and Earth History: The New Uniformitarianism.

Edited by W.A. Berggren and John A. Van Couvering.

Princeton University Press: 1984. Pp. 464. Hbk \$65, £60.10; pbk \$19.50, £13.90.

WHILE generations of geologists have been reared on Charles Lyell's doctrine of uniformitarianism, few have had an adequate appreciation of just what was meant by the term, or of why the issue was so contentious during much of his lifetime. His opponents, collectively dubbed "catastrophists" by one of them, William Whewell, have not had a fair press partly because Lyell's own version of the history of geological thought has been so widely and uncritically accepted. Certainly the uniformitarian doctrine is not satisfactorily rendered by the trite phrase taught to every geological student, "the present is the key to the past", because this actualistic principle was accepted by catastrophists and uniformitarians alike. What was more significantly at issue was whether, as Lyell believed, the Earth was essentially in a condition of steady state, or whether there had been progressive changes through time, with infrequent cataclysmic interruptions to a normally rather quiescent planet.

In recent years there has been considerable re-evaluation of the work of the leading early-nineteenth-century geologists, and a reaction against Lyell has set in. Lyell of course lost the argument about progressionism in the biosphere well before the end of the century, ironically because of the success of the theory of evolution propounded by his most ardent uniformitarian supporter, Charles Darwin, but the situation with regard to the inanimate world has been less clear-cut. It now appears that Lyell's steady-state Earth model is not at all a bad approximation for the past 2,500 Myr, provided that the cyclic changes inferred are considered as gross entities — the world of the Archaean was, after all, a closed book to nineteenth-century geologists. On the other hand there has been a resurgence of interest in "catastrophes" viewed as short-lived, comparatively violent interruptions to longer-lived episodes of relative quiescence, in fields as diverse as geomorphology, sedimentology, volcanology, tectonics and evolutionary palaeontology.

This volume contains 18 contributions to two symposia held in 1977, one at Woods Hole on the current state of uniformitarianism, and the other at the University of Kansas on the subject of the Cretaceous-Tertiary boundary. The deplorable lapse of time between the holding of the symposia and publication of the book has meant that a number of the articles seem badly dated, such that a few authors might feel embarrassed to see their work in print at this late

stage. Fortunately some chapters, especially those with substantial palaeontological documentation, are still well worth reading.

The longest and most interesting articles stemming from the Woods Hole meeting are the historical and philosophical essays by Gould and Benson. Both authors recognize that catastrophists such as Cuvier were the true empiricists of the day, interpreting the stratigraphic record as it appeared, for instance in the abruptly changing succession of fossil faunas, and that Lyell introduced confusion into the argument. Gould maintains that Lyell's gradualism, which so deeply influenced Darwin, was ideological in origin, imposing on the natural world a view derived from contemporary liberal politics, as opposed to the alternative Marxist view of history, propounded a few decades later, which claimed periodic revolutionary interruptions to a stable order.

It is always difficult to find firm supporting evidence for arguments postulating an influence by the contemporary *Zeitgeist*, because by its very nature the influence might be unconscious, but one can propose a more likely alternative. A link can be traced between Lyell's *Principles of Geology* and the earlier writings of James Hutton, though this is not acknowledged by Lyell himself:

In examining things present, we have data from which to reason with regard to what has been . . . Therefore, upon the supposition that the operations of nature are equable and steady, we find . . . a means of concluding a certain portion of time to have necessarily lapsed, in the production of those events of which we see the effects [J. Hutton in *Theory of the Earth*; my italics].

Thus the assumption of "equable and steady operations of nature" could be used as a heuristic principle to evaluate the passage of time, and it seems to me that here we have the key to Lyell's doctrine. If any politics were involved they were the politics of a scheming and ambitious scientist who failed to give acknowledgement where it was due and unfairly denigrated the work of his opponents.

The proceedings of the Kansas symposium include two general articles by Newell and Fischer. Newell reviews mass extinctions in their historical context, referring to the pioneer work of Cuvier and Brongniart in the Paris Basin, and looks to their cause in the chance coincidence of multiple factors such as changing habitat area, climate, and air and water chemistry. Fischer puts forward a stimulating model

involving two Phanerozoic super-cycles in which the Earth oscillates between an *Icehouse* and a *Greenhouse* state. The former is characterized by low sea-level, cold ice-capped poles and strong ocean currents, the latter by high sea-level, equability and widespread oceanic anoxia. A prime link is seen with plate tectonics. As continents split up, ocean ridges increase in volume and displace water over the continents. As a result of greater volcanicity producing more carbon dioxide and reduced land area limiting the amount of carbon dioxide lost by rock weathering, the atmospheric greenhouse effect causes a global rise in temperature.

In the light of the enormous interest generated over the "catastrophe" at the Cretaceous-Tertiary boundary, the articles of greatest value may well be the reviews of changes in various animal and plant groups across the boundary. These have been updated to 1980 and hence contain mention of the Alvarez hypothesis. Kauffman points out that extinctions in the marine realm were graded over a 1–5 Myr period and are primarily the result of general environmental deterioration produced by a fall in sea-level and water temperature. The extinction pattern is thought, however, to have probably been enhanced by some extraterrestrial event near the terminal phase, which served as "the last straw". Neither Hickey nor Tschudy find any evidence in the plant record of a terminal Cretaceous catastrophe consistent with the Alvarez dust-cloud scenario, as opposed to a more gradual increase in extinction rate caused by climatic deterioration. Similarly, Archibald and Clemens state that the turnover in terrestrial mammals was not catastrophic, indeed was no more marked than usually seen between any given succession of North American land-mammal ages. Russell, however, finds no evidence of a decline in dinosaur diversity before the end of the Cretaceous, and therefore feels that it is not possible to eliminate extraterrestrial hypotheses from consideration for this group at least.

Although the Alvarez team and others have recently found iridium anomalies at many more Cretaceous-Tertiary boundary localities, serious doubts and difficulties persist about extraterrestrially-induced catastrophes, especially as a *general* explanation for mass extinctions. The issue is likely to remain unresolved for a considerable time yet, and much cooperative analysis is called for between teams of specialists working on the Phanerozoic as a whole. Lyell accused his catastrophist opponents of trying to cut, rather than patiently unravel, the Gordian Knot. He would no doubt shake his head sadly at those modern catastrophists who favour a quick fix for a long-standing problem. □

A. Hallam is Lapworth Professor of Geology at the University of Birmingham, UK.

Changing fields

D.H. Tarling

The Earth's Magnetic Field.

By Ronald T. Merrill and Michael W. McElhinny.

Academic: 1983. Pp.401. Hbk £39.50, \$67.50; pbk £17.50, \$30.

Reversals of the Earth's Magnetic Field.

By J.A. Jacobs.

Adam Hilger, Bristol and Accord, Massachusetts: 1984. Pp.230. £19.95, \$35.

DURING the past decade there has been a dramatic increase in our knowledge of the behaviour and origin of the Earth's magnetic field. This revolution, if so it can be called, stems from (i) the greater information on the present field, available from satellite monitoring, (ii) on its past behaviour, as provided by palaeomagnetic studies, and (iii) from improved concepts of the nature of magnetohydrodynamic processes. It is no longer necessary to avoid questions of the origin of the geomagnetic field by simply stating that it is dynamo, and is now possible to discuss how that dynamo may operate and how it is driven. The problem has not been solved, but at least certain constraints are now accepted and we can pose hypotheses that are testable by further observations.

Merrill and McElhinny make it their aim to combine accounts of advances in observatory studies with those in palaeomagnetism and theoretical dynamo concepts in order to allow geomagnetic theoreticians to speak to palaeomagnetists, and vice versa. They have certainly provided a readable text; although some mathematical background is required for the more theoretical chapters, it is not beyond that which could be expected from a final-year geophysics undergraduate. The real surprise, given the book's title, is the lack of attention which is paid to satellite observations and to aeronomy. I would also have liked to have seen more critical appraisal of such things as the evidence for and significance of the frequency of polarity changes in the past, or the evidence for possible asymmetric polarity changes in the Precambrian — most such discussion does not go beyond that previously reviewed in other books. However, such a comment should be taken as a compliment about the generally excellent level of reviews presented.

The production is good, and the book is well indexed and, with some exceptions, well referenced — although one wonders why the historical references are not included. Mention of one reference published in June 1984, a week after receipt of the book for review, also contrasts starkly with the omission of any reference to Thellier's archaeomagnetic summary paper of 1981.

Jacobs's book deals with only one aspect of the geomagnetic field — its ability to

reverse — but this is, perhaps, the most important contribution that has and will still come from palaeomagnetic studies. The style is, as always, highly readable and the mathematical content is much less than in Merrill and McElhinny. Although Jacobs devotes the whole book to this one aspect — plus a few asides into climatic effects — the textual space is essentially similar as his book is only half as thick. There is, however, much greater and thorough discussion of the actual observations, Merrill and McElhinny being more concerned with consequences. Jacobs, for example, discusses the petrology: polarity correlations which are dismissed as a "data artefact" by Merrill and McElhinny. While such an assessment is probably correct, the argument is of interest. Another of R.L. Wilson's observations — the far-sided nature of the time-averaged palaeomagnetic pole — is similarly given greater consideration by Jacobs.

It is difficult to choose between these two books. For a graduate or postgraduate course on "pure" geomagnetism I would, in any case, marginally prefer the better balance towards aeronomy provided by W.D. Parkinson's *Introduction to Geomagnetism* (Scottish Academic Press/Elsevier Scientific, 1983). All three books really need to be in any geophysics library. Pedagogically I would be happy for any student to read either of these newer contributions. Both give helpful accounts of the solid-Earth aspects of geomagnetism and, to some extent, are complementary to one another — Jacobs providing a clearer general appraisal of the present data and Merrill and McElhinny giving a better appreciation of the implications of those data to geomagnetic theories. □

D.H. Tarling is Reader in Palaeomagnetism in the Department of Geophysics and Planetary Physics, University of Newcastle upon Tyne.

Pieces of a life

William Cooper

Random Variables.

By Lord Rothschild.

Collins: 1984. Pp.238. £12.50.

"Perhaps I can deal with your questions by telling you why we invited you to speak. You have had a rather unusual life, more like someone two hundred years ago. You have changed your job so often."

"25 years in a laboratory at Cambridge studying the biophysics of reproduction", I protested.

"Yes, but you were for 10 years the Research Director of Shell, and then you became a sort of civil servant. And you're a banker."

"Banker," I protested again: "Never."

"But you are called Rothschild."

Stunned, I was not quick enough to remind Colonel B. that the London telephone directory contains Rothschilds who are dentists, accountants, doctors and one who simply calls him or herself Trading Company.

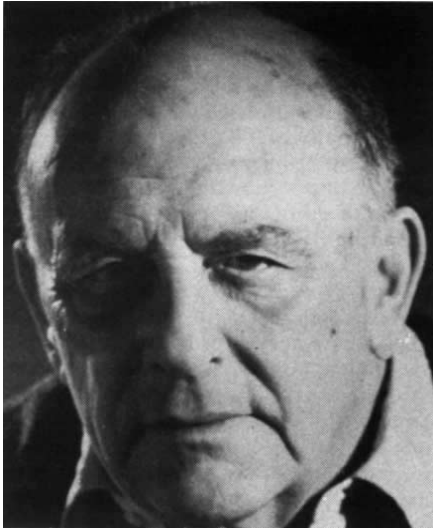
Do you pick up something of an authorial tone of voice in that extract? It is the tone of voice binding together this collection of pieces which Lord Rothschild presents under the title of *Random Variables*, artlessly recognizing in his preface that there is "a remote chance that a mathematician might be taken in by the title". A mathematician, quite possibly several readers of *Nature*; and, as a matter of fact, the present reviewer — thus disclosing my sole cause for any disappointment in the book: I was rather hoping for some amusingly far-fetched metaphor to make itself apparent. But no. "An abstract concept in mathematics", says Lord Rothschild, "not a subject for discussion here, even if the common reader were interested". So much for the likes of me! It's an entrancing book and I enjoyed every minute of it.

The eponymous variables consist of recollected episodes in the author's life, personal encounters that have caught his fancy, a speech or lecture here and there, comments on certain bits of government; plus two extended and fascinating pieces of family history. There is no common thread, other than the audible expression of the author's personality — clever, inquisitive, shrewd, high-spirited, and, no matter how deeply he is concerned, instinct with a fine carelessness that to me is most attractive. "This book, unlike that of my cousin Baron Guy de Rothschild", he says, "contains little or nothing about the *signes extérieures de la richesse*". What about *signes intérieures*? one feels inclined to ask. A wonderfully fine carelessness . . .

I suppose one would have to credit Colonel B. with an equally if antithetically fine carelessness; when he says a sort of civil servant, he is referring to Lord Rothschild's having been the first head (appointed by Mr Edward Heath, then Prime Minister) of the Government's Think Tank. Two of Lord Rothschild's variables are about the Think Tank, and they include personal encounters with Mr Heath and Sir Burke Trend and Sir William Armstrong, in which Lord Rothschild displays a novelist's knack of showing his characters being *themselves* — which aren't necessarily the selves they think they are projecting. Mr Heath dutifully makes statements which in effect carry the conversation forward not one iota. Sir Burke and Sir William hover in the background, their antennae anguishedly sensitive to the presence, in Lord Rothschild, of someone who may *rock the boat*. "Victor, if forced by circumstances to answer", Sir Burke says to him afterwards, "you may say that the Central Policy Review Staff is taking an interest in such-and-such a subject. That is as far as you may go. You may not say that you are

writing a report on any subject nor that you have written one".

I may not linger in detail on the remaining 26 variables, such as Lord Rothschild's ideas about the use and beauty of Royal Commissions (both liable to be nil); and a reported Court case in which, as plaintiff, he finds himself in Kafkaesque entanglement with LawSpeak. It is the chapters of family history which are the major pieces; guaranteed, I should have thought, to fascinate any reader by their combination of dramatic high affairs, huge sums of money and a stunning intelligence system. The first is about N.M. Rothschild, the founder of the family



Lord Rothschild — "a fine carelessness".

fortunes in Britain, who was supposed to have been present at the Battle of Waterloo, to have correctly observed which way the battle was going and to have returned instantly to London — there to buy up all the consols he could lay his hands on while they stood at a low price because of the city's pessimism about the result of the battle, and then to sell them next day when the price shot up at the news of victory. Lord Rothschild finds evidence that NM was not at Waterloo, and so the historical conundrum is set: How was the good news instantly brought from Waterloo to NM? And what happened then?

The second chapter, longer and more complex, is about NM's son, Baron Lionel de Rothschild, who put up the money for the British purchase, in competition with the French, of the bankrupt Khedive's shares in the newly-completed Suez Canal — outcome of Baron Lionel's loyal friendship with Disraeli, which enabled Disraeli to get advance intelligence from him about the Khedive's plans and then to procure from him the necessary funds with greater secrecy and speed than could have been possible otherwise. Remarkable chapters in a book that doesn't vary from being informative, amusing and enjoyable! □

William Cooper is the pseudonym of the novelist Harry Hoff. His most recent book is Scenes from Later Life (Macmillan, London, 1984).

Search for cognitive science

John Fox

Conceptual Structures: Information Processing in Mind and Machine.

By John F. Sowa.

Addison-Wesley: 1984. Pp.481.

\$28.95, £17.95.

ARTIFICIAL intelligence has produced the latest new technology to enter popular awareness. "Expert systems" are advanced computer systems which solve problems that were previously thought to be the preserve of human beings; not just any human being, but experts. Artificial intelligence (AI) techniques are used to emulate (to a first approximation) the problem-solving and decision-making of the expert human mind. Systems for medical diagnosis, geological data interpretation, semiconductor chip design, fertilizer selection, tax advice and so on have all been successfully developed to some degree, and the potential range of this new technology is enormous — even the emulation of exotica such as the military, scientific or political mind no longer seems entirely absurd.

In the middle of all this enthusiasm for "knowledge engineering" it is easy to forget that artificial intelligence is not new, and that it did not originally set out to be an engineering discipline but a science. The early workers defined their goal to be an understanding of the principles that underly intelligent behaviour in man, animals and (their special contribution) machines.

Artificial intelligence emerged in its modern shape when computers became available during the mid-1960s to scientists of many disciplines. Intelligence was seen as the capacity to manipulate symbols and, consequently, it could be understood in terms of symbol-manipulating models run on computers. Since the very early work on problem-solving and theorem-proving, researchers in AI have gone on to model the processes of visual perception, the understanding of language and speech, learning and the acquisition of intellectual skills, and even the formation of beliefs, social attitudes and the development of psychiatric conditions.

These subjects somewhat resemble those which have traditionally been the province of psychology. So what were psychologists doing during this period? Partly due to its own successes, and partly due to the influence of AI concepts, "cognitive psychology" was also developing rapidly in such fields as perception, memory, decision-making and so on. The relationship between the two communities was usually fraternal. The reference sections of psychological papers often acknowledged the contributions of AI, and vice versa.

Often, however, the acknowledgements were superficial only, and discussions were

peppered with critical exchanges. Programmers in AI may be vigorous and inventive, say many cognitive psychologists, but they are not *scientists* with the normal discipline for careful, systematic enquiry that the term implies. The counter is that psychologists are obsessed with methodology. They miss the point, and the insights, of AI demonstrations of what is possible with symbol-processing mechanisms.

Over the past ten years there have been a few, but generally less than successful attempts at synthesis, although a name has been invented, "cognitive science", which is supposed to cover both scientific AI and cognitive psychology. Unfortunately most cognitive scientists are only well informed about one of the branches. They are also trained very differently and their books and journals show it. It is against this background that John Sowa's book appears.

At first sight the book looks as though it might be something special indeed. The chapter headings cover a range of core ideas, including the philosophical basis of AI, reasoning and computation, language, psychological evidence and — above all — knowledge. Furthermore the preface declares that the book "combines AI and cognitive science approaches. . . . In combining insights from each of the separate fields, the book gives a unified view of knowledge representation".

Sowa applies a general plan to each chapter wherever possible, the first few pages being background material containing wide-ranging discussion. These sections usually summarize psychological research and findings (in perception, language and so on) before going into more detailed, technical presentations on the same subject where the emphasis is on computation and formalization. Sowa maintains an attractive, readable style throughout the book, studded with illuminating illustrations of his points from a remarkable variety of fields ranging from optical illusions perceived by the peasants of Uzbekistan to the structure of early Greek poetry.

This has the feeling of a mature work, a long time in the making and reflecting years of serious thought. It is surprising therefore that the book includes chapters on recent topics such as knowledge engineering as well as classical ones. This is one of its strengths as Sowa is able to cut through

New editions

● *Microbiological Methods* (5th Edn), by C.H. Collins and Patricia M. Lyne. Publisher is Butterworths, price is pbk £18, \$49.95.

● *The Biochemical Mode of Action of Pesticides* (2nd Edn), by J.R. Corbett, K. Wright and A.C. Baillie. Publisher is Academic Press, price is £35, \$59.

● *The Sulphate-reducing Bacteria* (2nd Edn), by J.R. Postgate. Publisher is Cambridge University Press, price is £20, \$39.50.

some of the AI jargon and relate the new ideas to established topics such as databases and query languages. The references show the same careful mix; they are comprehensive but up to date throughout.

The most obvious manifestation of the book's maturity, apart from style and balance, is that Sowa manages to convince the reader that most of his diverse material can be brought together under a single theoretical idea, the "conceptual graph". This is a formal, symbolic structure which can be manipulated by a computer, and which can represent many types of human knowledge. By the end of the book one is convinced that the concept may be sufficient to support machine-reasoning, language (syntax and semantics), decision-making, learning and (just possibly) creativity. Those familiar with AI will find many familiar themes, but the integration of subjects and the extensive investigation of the formal properties of symbolic structures is impressive.

I am less sure about Sowa's discussions of the human equivalents of these machine processes. Certainly his knowledge of and respect for the work of psychologists is refreshing. There is even a hint of affection for people, which I often miss in the machine-obsessed books of the AI world, however good they may be technically. But while the psychological material broadens and enriches the book, to say that the fields are "unified" is a little strong. More significantly, however, is that the attempt at "unification" is really an attempt on the terms of AI, not those of cognitive psychology. The links which are made between the performance of programs and observations made by psychologists can be somewhat superficial and unconvincing. They appear to be included to show us that programs can show interesting features of intelligence, not that they have truly human characteristics of intelligence. There is little effort to consider whether the key concept of the conceptual graph can explain in any detail the things that the human mind does not do, as well as things it can do. One suspects that — like other AI ideas — the conceptual graph is powerful enough to encompass many bad theories of mind as well as good ones.

Conceptual Structures is not the missing classic, the book that really achieves a unification of the two fields, but it is one of the best attempts that I have seen. Perhaps the AI traditions of imagination and vigorous inventiveness are still too far from the scientific tradition of systematic experimentation respected by cognitive psychologists. The book is nevertheless instructive and entertaining, detailed but economical. It would be on my list of six AI books to take to a desert island, and any student of cognitive science will find it very rewarding. □

John Fox is Head of the Biomedical Computing Unit at the Imperial Cancer Research Fund Laboratories, London.

Unhealthy business

Robert Ubell

The Medical Industrial Complex.

By Stanley Wohl.

Harmony Books, New York: 1984.
Pp.218. \$14.95.

IN his farewell address to the nation, President Dwight D. Eisenhower warned of the ominous encroachment of the "military-industrial complex" into American life. In 1980, Dr Arnold Relman, editor of *The New England Journal of Medicine*, raised another spectre, the threat of the "medical-industrial complex". Relman had noticed that profit-making groups had suddenly assumed ownership of a great network of hospitals, ambulatory surgical centres, free-standing emergency rooms, diagnostic laboratories, alcoholic treatment facilities and a range of other health-care services.

In little more than a generation, corporations have leaped from owning a handful to controlling about a third of the country's community hospitals, mostly in affluent suburbs. Today, for-profit systems own nearly 400,000 hospital beds as part of a \$118 billion industry, which includes the vast clinical supply business — pharmaceuticals, medical technology, food services and so on. One dollar in three spent on health care in the United States goes to corporations. Relman rightly feared that the result would be a return to "two-tier" care for Americans — one for those who can pay and another, far less adequate and less accessible, for the poor.

Adopting Relman's apt phrase as the title for his book, Stanley Wohl, a physician on the staff of the Stanford Medical Center and president of his own medical research and service company, offers us an exposé of the medical system. The dust jacket tells us that Wohl's book "picks up" from Paul Starr's Pulitzer-Prize-winning work, *The Social Transformation of American Medicine* (for review see *Nature* 304, 667; 1983). But Wohl provides neither Starr's historical perspective and keen analytical skill, nor does he go beyond Relman's original observations.

Wohl re-states what others have already concluded: the introduction of the corporation as a provider of health-care is likely to exacerbate, rather than relieve, the more intractable features of the system. To boost profits, corporations will further inflate the price of health care, despite claims that for-profit networks are best at containing costs through rational organization, economies of scale and tougher management. Recent studies, comparing "investor-owned" with non-profit hospitals, reveal that, on average, private chains charge more. Wohl quotes the head of American Federation of Hospitals, the business association of the private chains, as saying, "You can't be both newer and

cheaper and we never said we're cheaper".

American medicine is already the most wasteful in the world, annually spending \$1,500 per capita on health care — much more than any other nation. Despite Wohl's claim that the United States delivers "the highest quality [care] on earth", a new survey carried out by *The Economist* shows otherwise. In neither life expectancy nor infant-mortality rates — the two key indicators of national health — does the United States score best. A number of other industrialized countries are out in front. For instance life expectancy in Japan is 77 years (75 years in the United States) and infant mortality 7 per 1,000 live births (12 in the United States). And while (for example) Britain spends almost three-quarters less than the United States on health — about \$400 per capita — Britain is only a year behind the United States in life expectancy and equal in infant-mortality rate.

Wohl hopes that the American system can be saved by preserving the autonomy and authority of the physicians themselves. When doctors are in control, Wohl believes, the national interest will be served. But, by and large, American doctors have not performed well at keeping costs down, nor do patients fare best when physicians retain complete control. American doctors have fought a century-long battle to keep medicine as a profitable product rather than a public trust, ultimately setting the stage for the corporate invasion. Now, ironically, doctors are being forced to become company men, working for wages. Turning once again to the doctors for help does not appear to be wise.

In his comedy-thriller *The Man Who Knew Too Much*, Alfred Hitchcock guessed that the "commodification" of health care was one of the nastier secrets in American medicine. In that movie, Jimmy Stewart plays a mid-western doctor on holiday in North Africa. Doris Day plays his wife.

DORIS DAY: *Do you know what's paying for this trip?*

JIMMY STEWART: *What?*

DORIS DAY: *Mrs Campbell's gallstones. And the pearls you bought me in Paris? Paid for by Johnny Smith's tonsils.*

JIMMY STEWART: *You know, I never thought of it that way.*

DORIS DAY: *What would they think of us, if they heard us back home?*

The folks back home, then as now, uncritically accept the principle that medicine is a commodity, to be bought and sold on the market, with profits going to controlling interests. That is why Wohl's proposals for reform — which include the corporate chains as a key ingredient in medical delivery — act much like a thin hospital sheet. As he pulls it up to cover his head, his toes stick out in the cold. □

Robert Ubell is a scientific and medical publisher and editor in New York.